EXAMINING THERMOKARST INITIATION WITH RANDOM FOREST MODELS

By

Rawser W. Spicer B.S.

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Computer Science

University of Alaska Fairbanks

May 2020

APPROVED:
W. Robert Bolton, Committee Chair
Orion Lawlor, Co-Chair
Glenn Chappell, Committee Member
Chris Hartman, Chair
Department of Computer Science

Abstract

This project examines thermokarst initiation through the application of random forest models. Thermokarst initiation marks the start of the formation of thermokarst features. Changes in landscape, due to the thermokarst process, can result in changes in wildlife habitat, as well as energy, carbon and water fluxes. Random forests are an ensemble learning technique that combines the results of many independent decision trees to create results that avoid the overfitting in regular decision trees. Random forests were trained against an existing thermokarst initiation model. Results showed that random forests were useful in this context. Random forest hyperparameters were also examined through a multiparameter sensitivity analysis.

Contents

1. Introduction

Arctic landscapes are changing at an increasing rate (Jorgenson, Shur and Pullman 2006) (Lara, et al. 2014). One process driving these changes is thermokarst. This is a process that occurs as permafrost and ground ice thaw due to warming surface temperatures (Jorgenson, Shur and Pullman 2006) (Lara, et al. 2014). Modeling these changes is important to understanding how arctic environments will react in the future which has direct applications in infrastructure improvement, and scientific understanding of the arctic. This research applies using random forests (RF) to model an existing method of thermokarst initiation (TKI) to identify regions where ground ice thaw may begin, the start of the thermokarst process. The sensitivity of random forest models to changes in input features, and changes in hyperparameters are examined.

Random forests are an ensemble of decision trees that have been used for classification and regression, and applications include weather forecasts, classification of remotely sensed data, and other ecological predictions (Breimen 2001) (Cutler, et al. 2007) (Herman and Schumacher 2018) (McGovern, et al. 2017) (Pal 2007) (Wang, et al. 2016) (Were, et al. 2015). Random forests use a set of input variables, called features, to find some output value, usually called labels. RF models are generally trained a on subset of known features, and labels. These models can also be used to determine how much each input features affects the decision process, and therefor how important they are in relation to other variables (Kanevskiy, et al. 2017). This importance ranking can be applied to the development of reduced order models. Reduced order models (ROMs) aim to create models that are less complex than existing models in order to increase speed. These models should have a small approximation error, conserve the properties of the original model, and be computationally efficient (Antoulas 2004).

Changing arctic landscapes affect vegetation, wildlife, hydrologic drainage patterns, and the carbon & energy regimes (Gandodamage, et al. 2014) (Jorgenson, Shur and Pullman 2006) (Jorgenson, et al. 2015) (Lara, et al. 2014) (Kanevskiy, et al. 2017) (Raynolds, et al. 2014). Understanding these changes is important for ecosystem, hydrology and permafrost sciences. Knowing were and how landscapes will change can inform wildlife managers in deciding what areas to protect as important habitats change, and development in the arctic region continue. As the landscape changes surface water distribution will change leading to potential changes in carbon, energy, and water fluxes . Finding how this will occur can increase our understanding of feedbacks to the climate system (Jorgenson, Shur and Pullman 2006) (Jorgenson, et al. 2015) (Lara, et al. 2014) (Liljedahl, et al. 2016).

This report examines background related to thermokarst, and random forests in section 2. Sections 3 through 5 look at the data and methods used to perform the

analysis. Section 6 looks at identifying a random forest to use, and how random forest models respond to changes in features, and hyperparameters. A final model is also presented. The remainder of the report discusses these results, and future improvements this process and applications of random forests to other problems.

2. Background

2.1. Permafrost

Permafrost is soil that is frozen (less than 0 °C) for two or more consecutive years, and occurs primarily in polar and alpine environments (Davis 2001) (Pollard 2018) (Rowley, et al. 2015). Permafrost has its greatest extent in areas where it has existed for thousands of years or longer. Permafrost generally occurs terrestrially, but also exists in the seabed in polar regions (Davis 2001) (Pollard 2018). Permafrost affects approximately 25% of the Earth's land area mostly in the arctic and sub-arctic (Pollard 2018) (Rowley, et al. 2015). Permafrost covers approximately 80% of Alaska, 50% of Canada, and 60% of Russia (Pollard 2018).

Terrestrial permafrost is classified into three groups: continuous, discontinuous, and sporadic. Continuous permafrost covers 90% or more of a specific landscape with mean annual soil temperatures around -8 °C. Discontinuous permafrost occupies 50-90% of landscapes in permafrost areas with mean soil temperatures around -5 °C. Sporadic permafrost covers less than 50% of these environments, and the annual soil temperatures near 0 °C (Pollard 2018) (Rowley, et al. 2015). Stable mean annual temperature is important to permafrost, as low heat flux is important for maintaining permafrost. Soil moisture, air temperature, snow cover, aspect, and elevation also influence the local heat fluxes. (Pollard 2018) (Rowley, et al. 2015)

Permafrost depth ranges from more than 1000m to only a few meters at its southern limits. The deepest known permafrost occurs in Siberia and is around 1400m thick (Davis 2001) (Pollard 2018) (Rowley, et al. 2015). The maximum depth of permafrost is limited by heat in the earth's mantle (Rowley, et al. 2015). Most permafrost is thousands to millions of years old. However, permafrost has formed more recently primarily in regions of discontinuous (Davis 2001) (Pollard 2018).

2.2. The Active Layer

During summer, as temperatures rise above 0 °C, the top layer of the soil column seasonally thaws. This seasonal thaw layer is known as the active layer (Davis 2001) (Pollard 2018) (Rowley, et al. 2015). The active layer depth ranges from a few centimeters in the far north to several meters in the discontinuous permafrost zone

(Davis 2001). The active layer buffers permafrost from warm summer temperatures (Pollard 2018). The conditions at the surface of the soil determine the influence the of air temperature on frozen soils. These conditions include slope, aspect, soil moisture, and snow cover (Rowley, et al. 2015). Warmer summer temperatures are increasing the active layer depth and increasing destabilization of near surface permafrost (Pollard 2018) (Rowley, et al. 2015).

2.3. Ground Ice

Ground ice is any ice in freezing or frozen soils and occurs extensively in permafrost zones (Rowley, et al. 2015). There are many types of ground ice ranging from massive ice deposits to small ice crystals in soil called pore ice. Ground ice can be ice that has formed in the soil or ice that has become covered as soil is deposited (Pollard 2018). Ice wedge arrays are a type of ground ice that occur in an estimated 10% of the permafrost in Alaska's arctic coastal plain (Davis 2001).

Ice wedge ground ice occurs as moisture seeps into cracks in permafrost. These wedges grow as the ice freezes and thaws allowing more water in. These wedges reach 10m in depth and range from 2-3m wide (Davis 2001). Networks of ice wedge features come together in distinctive polygonal patterns called ice wedge polygons that range in size from 8 to 18 m across (Davis 2001) (Rowley, et al. 2015). Ice wedges, along with other ground ice features, are prone to completely melt as temperatures warm (Pollard 2018). This leaves distinctive depressions in landscapes through a process called thermokarst (Davis 2001).

2.4. Thermokarst

Thermokarst describes the melting of ground ice, and the resulting landscape features due to subsidence of ground surface(Farquharson, et al. 2016) (Kanevskiy, et al. 2017). This process is caused by disturbances to the ground surface and accelerated by increasing thaw in permafrost (Farquharson, et al. 2016) (Jorgenson, Shur and Pullman 2006) (Jorgenson, et al. 2015) (Kanevskiy, et al. 2017) (Raynolds, et al. 2014). One type of thermokarst is polygonal patterned ground formed by the melting of ice wedge features. The melting of these wedges can eventually lead to the formations of ponds and lakes (Jorgenson, Shur and Pullman 2006) (Kanevskiy, et al. 2017) (Liljedahl, et al. 2016) (Raynolds, et al. 2014). Simulating ice wedge polygons transitions, and their effects on landscapes are the main goal of the Alaska Thermokarst Model(ATM) (Bolton, et al. 2018).

2.5. Alaska Thermokarst Model

The Alaska Thermokarst model aims to predict how landscapes change in response to the thermokarst process. It is an element-based state and transition model that tracks the percentage of each landscape cohort in all pixels and changes these in response to thermokarst. The temporal resolution of the ATM is one year. Landscape cohorts describe the characteristics of the landscape. These include meadows, different types of polygonal patterned ground, lakes and ponds (Bolton, et al. 2018). The ATM is currently being developed.

The frame-based process of the ATM asks four main questions at each timestep to determine landscape change (Figure 1). One question, has thermokarst been initiated by an extreme event (question 2), is the primary research presented in this report. The thermokarst initiation model presented here can be considered a subprocess with in the ATM (Bolton, et al. 2018). The current method for modeling TKI is further described in section 5.1.

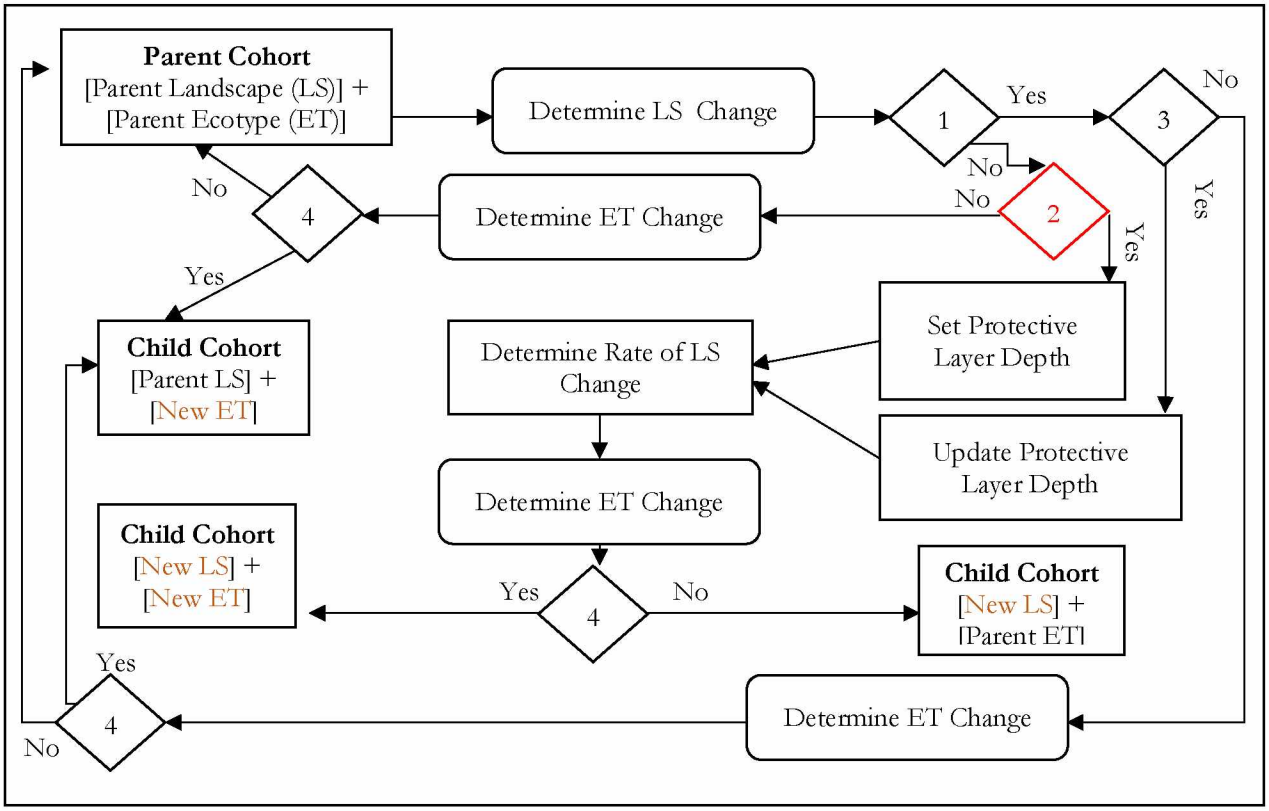

Figure 1: Flow Chart of ATM Logic. 1. Is thermokarst Active? 2. Has thermokarst been initiated by an extreme event? 3. Is Active layer > protective layer? 4 Does climate support new ecotype?

2.6. Random Forests

Random forest models are an ensemble machine learning technique that combines the results of many decision trees. Random forest models can be used as classification or regression tools (Breimen 2001) (McGovern, et al. 2017). Each tree in a random forest is constructed using a randomized subset of the training data, and random subset of questions at each node. Final results are calculated by taking the mode (for classification) or mean (for regression) of all the trees correcting overfitting that can occur with a single decision tree (Breimen 2001) (McGovern, et al. 2017). The mechanics of each concept used in random forests are described in the following sections.

2.6.1. Decision Trees

Decision trees are a network of decision nodes and leaf nodes. Each decision node has exactly two child nodes which may be either another decision node or a leaf node. The split at each decision node is based on comparing the value of a single input feature. Predictions travel left or right based on whether this comparison is true or false. For a prediction, the tree is traversed for a set of inputs through each node, starting at the root, until a leaf node is reached. Leaf nodes contain the final result of the prediction. For classification this is a label, while for regression this is a numerical value (Herman and Schumacher 2018) (McGovern, et al. 2017).

Figure 2 shows an example decision tree based off the TKI random forest models presented in this paper. The tree is three levels deep which results in eight leaf nodes. Each decision node shows the question asked (for example the root asks is thawing degree days (TDD) <= 1239.1), mean square error for the split, the percentage of samples left for further nodes, and the value that the node would return. The leaf nodes contain the same information minus the question.

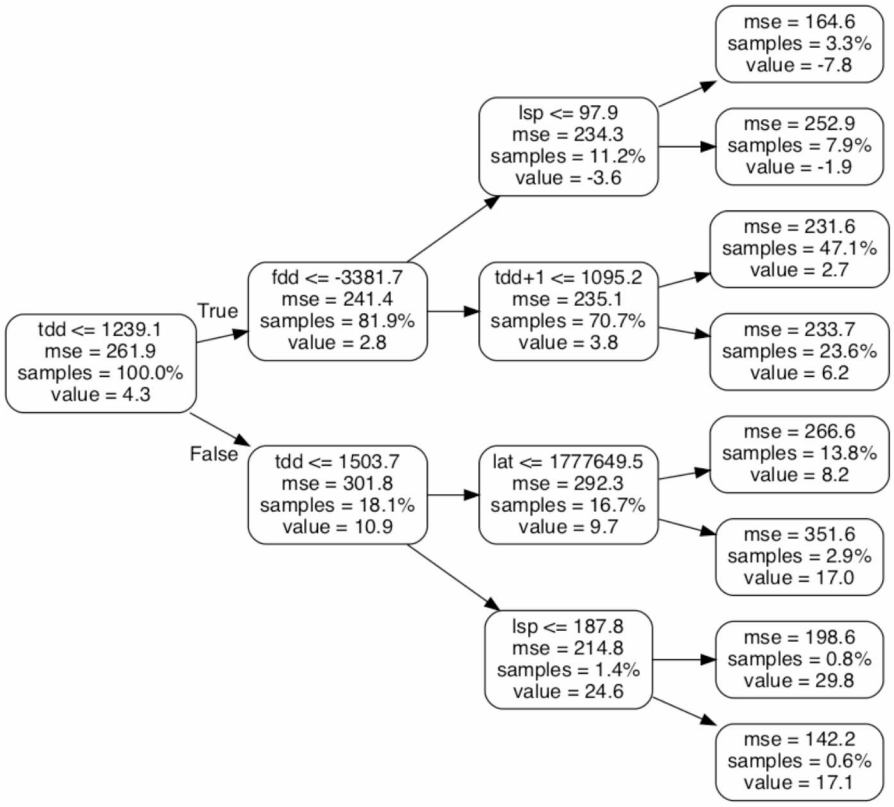

Figure 2: Example Decision Tree

2.6.2. Bagging

Bootstrap aggregating, or bagging, is a method used to randomize data used in statistics. In general, bootstrapping refers to selecting a random subset of samples from a given set of data with replacement (Breimen 2001) (Hastie, Tibshirani and Jerome 2017) (Herman and Schumacher 2018) (McGovern, et al. 2017). Giving a training set, size *n*, bagging builds modified training data, also size *n*, for each tree by selecting *n* samples with replacement from the original training data. Sampling with replacement selects from the full original data for each new, so the same item may appear more than once. This method is called tree bagging. Feature bagging can also occur, and only selects a subset of features at each decision node. Using these bagging techniques results in largely uncorrelated trees (Breimen 2001) (Hastie, Tibshirani and Jerome 2017) (Herman and Schumacher 2018) (McGovern, et al. 2017).

2.6.3. Applications

Random forest models have been used as precipitation forecasting models. Using 11 years of data from NOAA's Second-Generation Global Ensemble Forecast System Reforecast, random forest models were created that could predict accumulated precipitation for 2 and 3 day periods for regions across the contiguous United States. It

was found that random forest models were capable of predicting extreme precipitation events over other methods examined (Herman and Schumacher 2018). RF techniques have been used to classify landscapes from remotely sensed data in ecological and agricultural fields (Cutler, et al. 2007) (Pal 2007). Additionally, RF regression is used to predict soil carbon and biomass concentrations (Wang, et al. 2016) (Were, et al. 2015). In many cases RF results are compared to, and found to be comparable to, or outperform, other machine learning methods like artificial neural networks (ANN), and support vector machines (SVM) (Cutler, et al. 2007) (Pal 2007) (Wang, et al. 2016) (Were, et al. 2015).

3. Study Area and Time Resolution

The study primary area consists of Alaska's Arctic Coastal Plain(ACP) and extends south of the Brooks Range (Area A in Figure 3). All data used is in the form of geotiff raster files to ensure consistency of the locations of each cell. The data have all been converted to a $1km^2$ scale. The areas dimensions 415 rows by 1096 columns. The secondary study area consists of the Seward Peninsula (SP). This area is considerably smaller, and is where the observations for the original TKI concepts are based. The second study area (Area B in Figure 3) has the same model resolution as the ACP data, and has a model domain of 270 row by 384 columns.

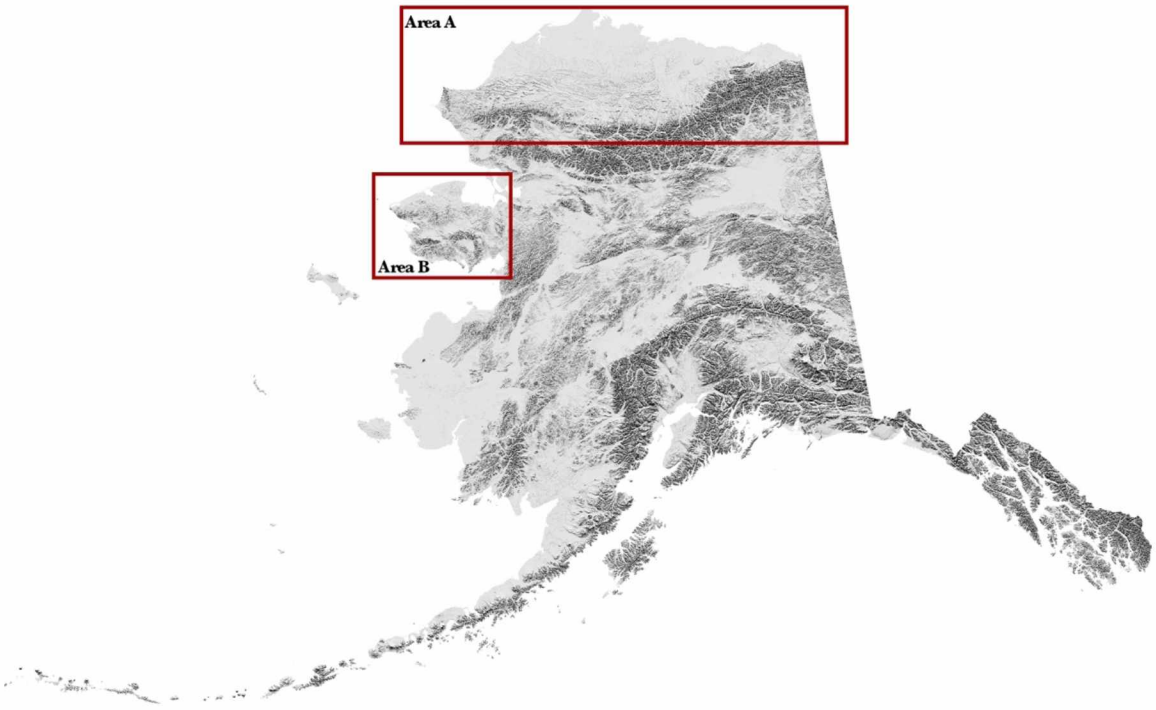

Figure 3: Study Areas. Area A: The Arctic Coastal Plain. Area B: The Seward Peninsula.

4. Data

The climatological data exists for a period of 115 years starting in 1901 and was derived from Scenarios Network for Alaska and Arctic Planning (SNAP) data (Scenarios Network for Alaska and Arctic Planing 2019). This is also the time period used to create the training data for the random forest models described later. For the random forest models the data is sub-sampled into 25, 50, and 75 percent portions by taking random raster cells from across the full 115 year period. For data were there is no change, or very little, from year to year like geolocation the rasters are assumed to be constant over time. The raster projection used is the NAD83 Alaska Albers equal-area conic projection (EPSG:3338) (spatialreference.org 2006).

4.1. Air Temperature

Air temperature is considered because it is a factor that controls active layer depth which indicates that ground ice experiences warming (Kanevskiy, et al. 2017). Air temperature data is used in a degree day format which represent heating or cooling. Here freezing Degree-days (FDD) represent winter temperatures while thawing degree days (TDD) represent summer temperatures. Degree-days are the sum of the departure of all daily average temperatures from a base value, in this case 0°C, for given period. The periods looked at here are summers and winters which were found by integrating a curve created from monthly average temperature data. The Degree-day data used was derived from SNAP Historical Monthly Temperature - 1 km CRU TS data which contain downscaled estimates of monthly mean temperatures raster data at a 1 km square resolution for 1901 to 2015 (Scenarios Network for Alaska and Arctic Planing 2019). Data for the ACP was clipped to the extents described in section 3.

To calculate FDD and TDD from the monthly data, a spline interpolation provided by Scipy was used. A univariate spline function was fit to the data, producing a periodic function with 2 roots per year, occurring at 0°C, representing the change from summer to winter or winter to summer. Counting the number of roots and comparing to the number of years times two was used to verify the correct number of roots. These number are required to be equal, and it was assumed that these roots occurred twice per year due to the smoothness of the curve. The spline function was integrated for all periods from root N to N+1 for all roots of the function in the given data range producing a set of alternating negative (FDD) and positive (TDD) values for each year. This process was done for each element in the model domain. For the few elements (less than 10 locations) where the method failed, FDD and TDD were determined by interpolating surrounding elements. FDD values were labeled according to the year where the winter starts. Examples of freezing degree day and thawing degree day data are shown in FiguresFigure 4 andFigure 5.

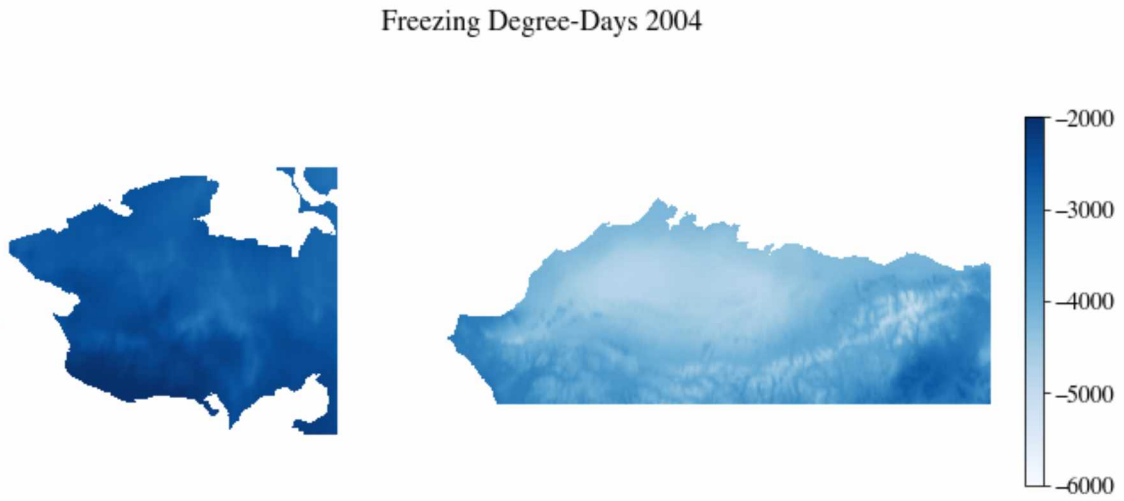

Figure 4: Freezing Degree Days for 2004 (Seward Peninsula Left, ACP Right)

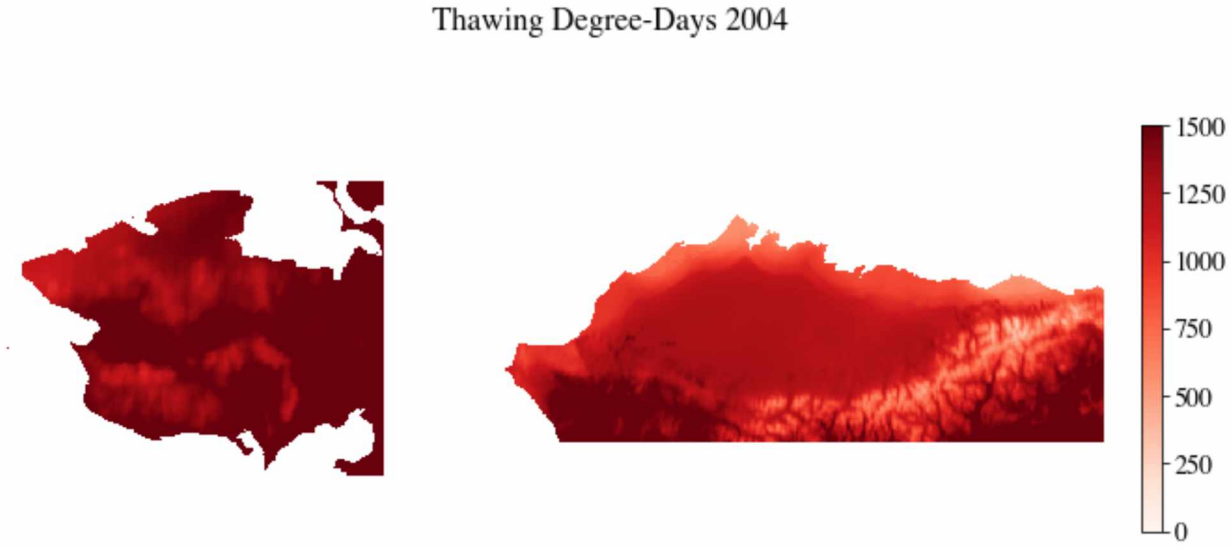

Figure 5: Thawing Degree Days 2004 (Seward Peninsula Left, ACP Right)

4.2. Precipitation

Precipitation, through both summer rain and winter snow, is another factor that effects active layer depth (Kanevskiy, et al. 2017). Precipitation data are in total millimeters and were derived from SNAP Historical Monthly Precipitation - 1 km CRU TS data which contain data downscaled to the same parameters as the SNAP air temperature data (Scenarios Network for Alaska and Arctic Planning 2019). The monthly data are summed into the desired time periods used in the TKI and random forest models. These periods are summarized in Table 1. Example precipitation figures are in appendix A.

Table 1: Summary of precipitation periods

Name	Months
Summer Precipitation (SP)	April – September
Late Summer Precipitation (LSP)	August & September
Early Winter Precipitation (EWP)	October & November
Full Winter Precipitation (FWP)	October – March

4.3. Location

Location data is derived from the raster data for FDD, though any georeferenced raster for the proper area of interest would work. The affine transform, which describes how to convert from pixel coordinates to world coordinates, was used to calculate the northing and easting in meters in NAD83 Alaska Albers equal-area conic projection (EPSG:3338), for the study area. Northing corresponds with latitude, and easting corresponds with longitude. Location data were then saved as raster files where the values represent northing and easting for the center of the model element.

4.4. Elevation, Slope, and Aspect

Elevation data is derived from the national elevation dataset 60 m hill shade product provided via the state of Alaska elevation data portal (DGGS n.d.). The 60 m data was rescaled to a 1 km^2 resolution. Slope and aspect were calculated from the 1 km^2 data using QGIS.

4.5. Summary of Features

Table 2 summarizes the data used as features in the initial Random Forest models. The table notes if the data are used in the original TKI model and if features are climate related or physical location based. The source of the data is also listed.

Table 2: Summary of Features

Name	**Used in Original TKI Model**	**Type**	**Source**
Freezing Degree Day (FDD)	Yes	Climate	Derived from SNAP monthly air temperature data
Thawing Degree Day - First Summer (TDD)	Yes	Climate	Derived from SNAP monthly air temperature data
Thawing Degree Day - Second Summer (TDD+1)	Yes	Climate	Derived from SNAP monthly air temperature data
Summer Precipitation - First Summer (SP)	No	Climate	Derived from SNAP monthly precipitation data
Late Summer Precipitation - First Summer (LSP)	No	Climate	Derived from SNAP monthly precipitation data
Early Winter Precipitation (EWP)	Yes	Climate	Derived from SNAP monthly precipitation data
Full Winter Precipitation (FWP)	Yes	Climate	Derived from SNAP monthly precipitation data
Summer Precipitation - second Summer (SP+1)	No	Climate	Derived from SNAP monthly precipitation data
Northing (Lat)	No	Physical	Derived from raster affine transform
Easting (Long)	No	Physical	Derived from raster affine transform
Elevation (Elev)	No	Physical	Derived from AK DGGS dataset
Slope	No	Physical	Derived from elevation
Aspect	No	Physical	Derived from elevation

5. Methods

5.1. Climate Priming Thermokarst Initiation Model

The original thermokarst initiation Model is based on the idea that consecutive, and sequential, extreme climate conditions can "prime" the local landscape for thermokarst initiation. High temperatures in summer and winter, increased snow cover, and extreme summer precipitation all lead to higher soil temperatures which can lead to increase in the active layer depth creating conditions that lead to thermokarst (Kanevskiy, et al. 2017). For example, the 2017-18 winter produced a large snow pack and extreme warm temperatures on the Seward Peninsula creating conditions in which the soil column did not completely freeze (Bolton, et al. 2018). Based on this idea the "climate priming" TKI model looks for years where the summer temperature, and the preceding summer and winter temperatures, and winter precipitation were greater than the average long term average. For each of these values the percent difference from the average was calculated. These presented differences were then averaged to find the TKI value. The climate priming TKI method is presented in Algorithm 1. The variables that were used for the TKI model presented here were TDD, EWP, FWP, FDD and TDD+1. Figure 6 shows examples of the results of the climate priming TKI for 1965 a regular year, and 2004 a year with high TKI values. Further examples are in appendix B.

1: Let Y be years from y_s to y_e
2: $y_{em} \in Y$ is the end of the period used in means
3: Let t be thermokarst predisposition value [0,1], or 1 when not applying predisposition
4: Let V be set of variables with values for $y \in Y$
5: Let M_v be the mean for the period $[y_s, y_{em}]$ for each v in V
6: Let PD_v be $\left\{\left(\left(V_y - M_v\right)/|M_v|\right) * 100 \text{ for } y \in Y\right\}$ for $v \in V$
7: Let TKI_y be $\frac{t}{n_V}\sum PD_{v_y}$

Algorithm 1: Mthod for finding TKI. Y = the range of years from a start year (y_s) to an end year for all pedictions (y_e). y_{em} = the final year for the mean period. t = thermokarst initiation value. V = the set of climate variables. M_v = the mean for a given variable from y_s to y_{em}. PD_v = the percent difference form M_v for each year in Y for each variable in V. TKI_y = the thermokarst initiation value

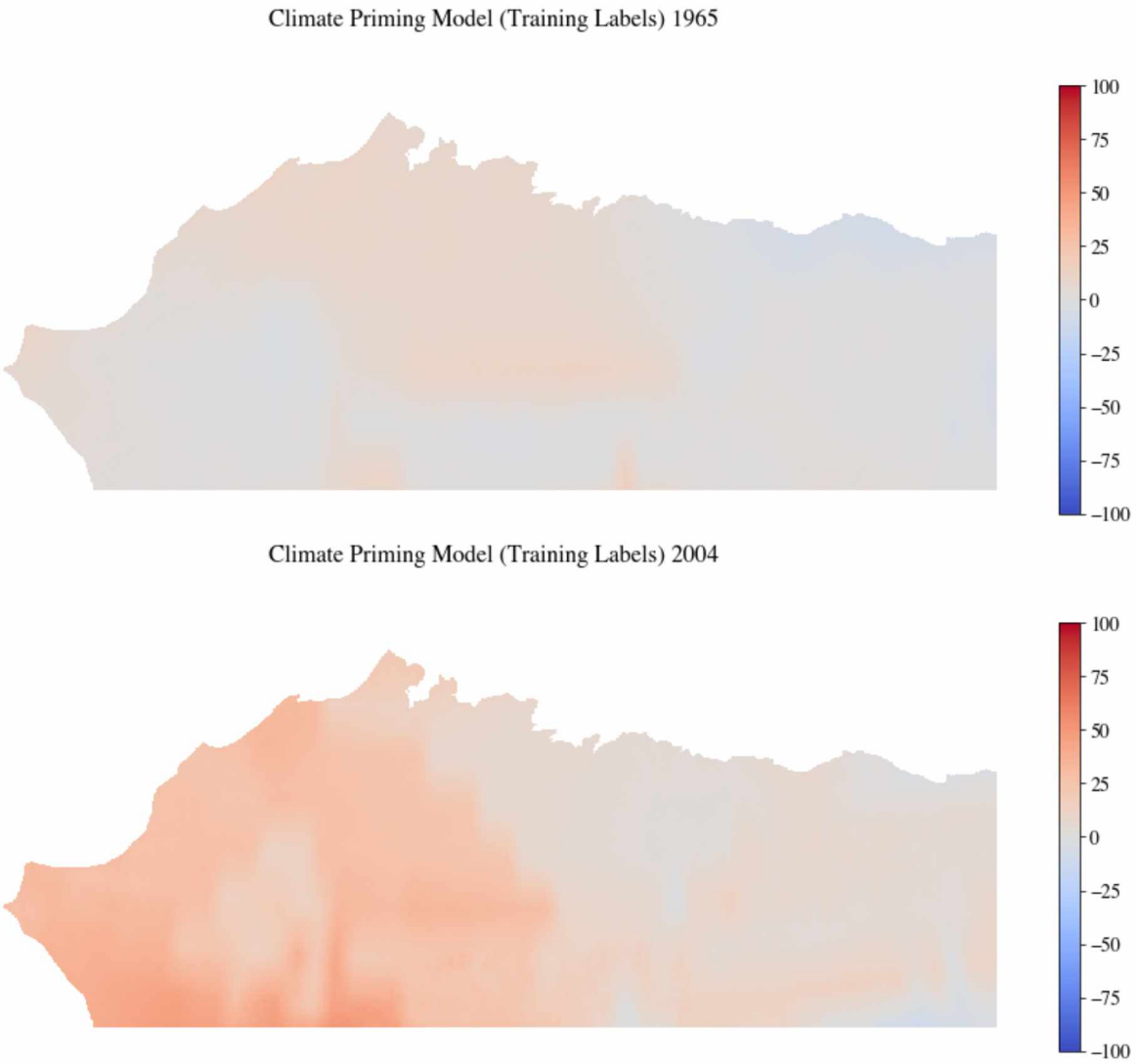

Figure 6: ACP TKI for 1965, a normal year, and 2004, an extreme year.

5.2. Software

Model implementation and analysis were done in Python (Version 3.7.3). The Conda package manager was used to create the Python environments used. It allows for the creation of isolated python environments that can be shared between systems (Conda n.d.).

The Random Forest regressor in Scikit-learn (version 0.20.3) was used to create the random forest models. This random forest implementation has many useful features including providing the feature importance, and ability to provide a decision path. There are also several hyperparameter available to control tree growth. If these parameters are unset their defaults create fully grown trees which can be very large (scikit-learn-developers n.d.).

5.3. Finding Baseline Hyperparameters

Hyperparameters are parameters to random forest models, and other machine learning methods, that describe how the forest is grown and set before training begins. RF models have several hyperparameters that can be adjusted to affect the accuracy of the model being trained. Determining what values of these hyperparameters optimizes the model results can be difficult. Scikit-learn provides access to several hyperparameters include the number of trees, the number of decision features at each node, and the maximum depth of each tree. Other parameters influence how the trees are split at each decision node (Pedregosa, et al. 2011). The Scikit-learn hyperparameters considered for this study are summarized in Table 3.

Table 3: Summary of Random Forest Hyperparameters

Hyperparameter	**Scikit-learn RF argument**	**Description**
Number of Estimators	n_estimators	Number of trees in the forest.
Maximum tree Depth	max_depth	Maximum depth of each tree.
Minimum samples to split a node	min_samples_split	Number of samples required to split a node.
Minimum samples per leaf	min_samples_leaf	Number of samples required at each leaf.
Number decision features	max_features	Maximum number of features considered at each split.
Maximum leaf nodes	max_leaf_nodes	Maximum leaf nodes of each tree.
Percent training data	Percent Training Data	Percentage of data used for training model

In order to determine which combination of hyperparameters create strong random forest models we looked at various combinations of parameters. The values of each hyperparameter considered are shown in Table 4. The combinations of these hyperparameters were examined in a brute force manner using the ACP region for training. Each variation of a hyperparameter was combined with each possibility of other the hyperparameters. Using this method resulted in 5,184 models to train.

Table 4: Summary of Hyperparameters

Hyperparameter	**Values**
Number of Estimators	10, 50, 100
Maximum tree Depth	12, 25, 60, 100
Minimum samples to split a node	2, 5, 10
Minimum samples per leaf	1, 2, 4, 8
Number decision features	AUTO, SQRT, LOG2
Maximum leaf nodes	1,000, 5,000, 10,000, 50,000
Percent Training Data	25, 50, 75

Training the 5,184 possible models was accomplished using four computing systems with varying capabilities, and existing workloads. They ran either Mac OSX, Debian, or Ubuntu. Table 5 summarizes the systems used. The Random Forest regressors were configured in a way that four parallel jobs were used to build trees in the forest. Statistics were collected on each of the models trained, but the models themselves were not saved at this time. The statistics recorded included: the time to train each model; the time to predict the values for a given year; the mean difference and variance; the mean absolute difference and variance, and the median.

The work done by each computer was tracked via a comma separated value (csv) file located in a git repository. This list initially contained the parameters to be used, a column to track the progress of the training, and empty columns for the desired statistics to be collected. After each model was trained, the csv data was updated with the model's statistics, and synced with the git repository. During this step, the next set of Random Forest parameters was also acquired, and steps are taken to avoid merge conflicts.

Table 5: Training Systems Summary

Computer Name	**OS**	**CPU**	**Memory**
Bristlecone	Ubuntu 18.04	3.2 GHz Intel Core i7 (12)	16 GB 1333 MHz DDR3
Finwhale	Debian 10	3.2 GHz Intel Core i7 (12)	16 GB 1333 MHz DDR3
Ocotal	Ubuntu 18.04	2.4 GHz Intel Xeon (16)	24 GB 1067 MHZ
Chickadee	Mac OSX 10.12	2.5 GHz Intel Core i7 (4)	16 GB 1600 MHz DDR3

The bulk training of the random forest models was run over the period of about a month before the training programs were externally stopped. A combined training time of 126 computing days was completed during this period. The order for training took the number of estimators into account before any other hyperparameter. Therefore all 10 and 50 estimator models were run during the testing period while only ~83% of all 100 estimator models were completed before the training period was halted. Figure 7 Figure 8 show how the model training times vary color coded by number of estimators. Timing data was largely inconsistent, and it was difficult to find how changing hyperparameters affected timing from this data. This could be due to the nature of the many changing hyperparameters, and the fact that the systems were not equally powerful because of this timing was largely not considered as a performance measure.

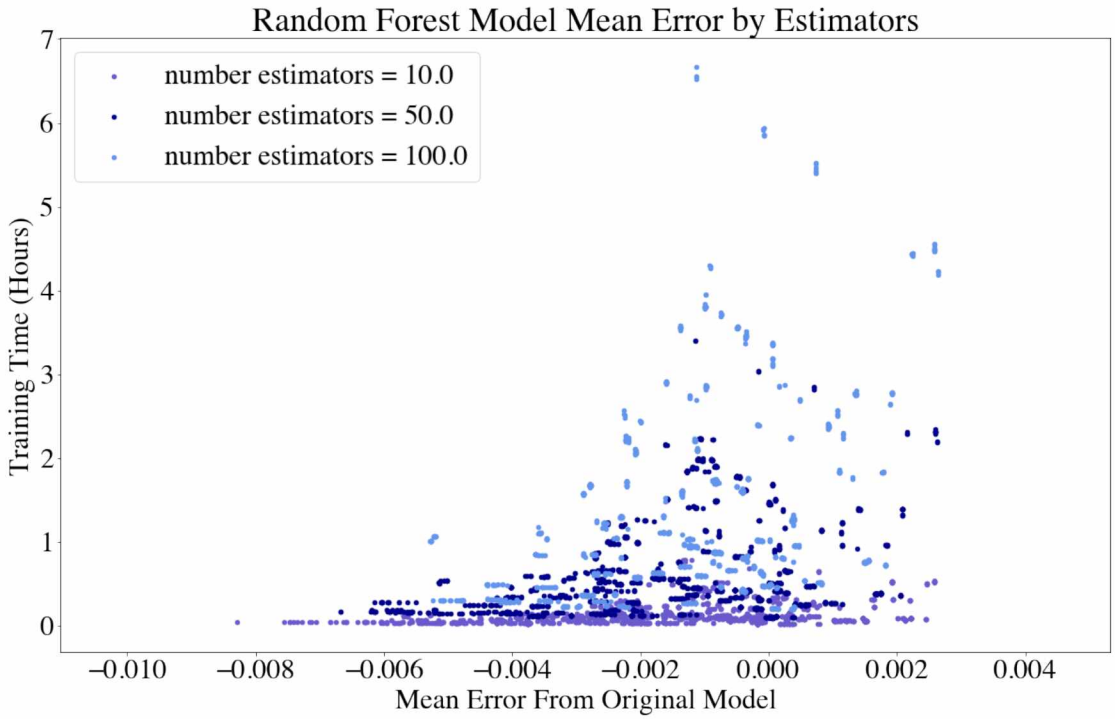

Figure 7: Mean Error. Notice that all mean error values remain near 0, and no clear timing relation is apparent in this plot

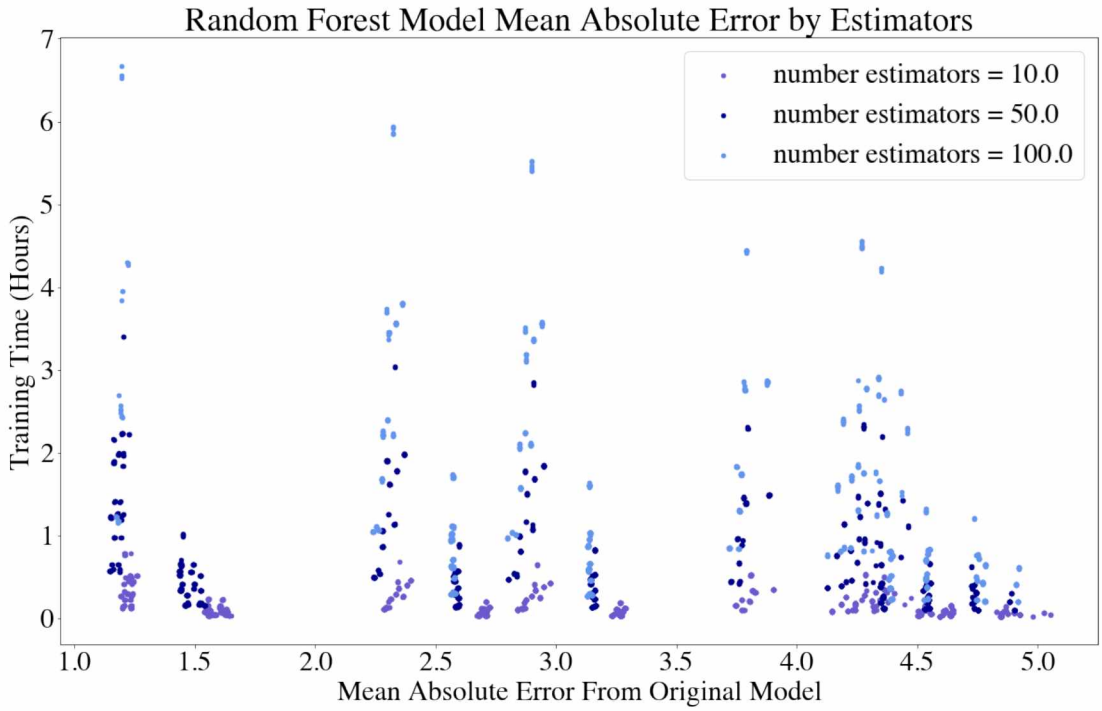

Figure 8: Mean Absolute Error. Notice All MAE values are greater than 0 giving a better indication of model accuracy.

5.4. Comparing Random Forest Models

Statistics that can be used to measure the accuracy of random forest models include the mean error (ME), the mean absolute error (MAE), and the coefficient of determination (R^2) for each model. The equations are shown in Equation 1, 2, and 3 where RF refers to the predictions of the random forest model and TKI refers to the "climate priming" thermokarst initiation model predictions which is considered the expected value in these comparisons .

$$ME = \frac{\sum_{i=1}^{n} RF_i - TKI_i}{n}$$

Equation 1: Mean Error; where ME = the mean error, RF_i = random forest prediction for an element; TKI_i = "climate priming" TKI model prediction for an element; n = number of elements.

$$MAE = \frac{\sum_{i=1}^{n} |RF_i - TKI_i|}{n}$$

Equation 2: Mean Absolute Error; where MAE = the mean absolute error, RF_i = random forest prediction for an element; TKI_i = "climate priming" TKI model prediction for an element; n = number of elements.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (TKI_i - RF_i)^2}{\sum_{i=1}^{n} (\mathrm{TKI_i} - \mathrm{TKI_{mean}})^2}$$

Equation 3: Coefficient of Determination; where R^2 = the coefficient of determination, RF_i = random forest prediction for an element; TKI_i = "climate priming" TKI model prediction for an element; n = number of elements.

Both ME, and MAE were used in the evaluation as ME is inadequate for showing how far from the TKI model predictions are. This is because many of the error values are above and below zero, the expected value, and when the average is taken the ME is always close to zero. The MAE corrects this by taking the absolute value of the differences before calculating the mean there for giving a measure of how far off the random forest model is predicting. Figures 4 and 5 show this phenomenon graphically for the models that were trained during the brute force testing.

The coefficient of determination, R^2, is another measure of how well observed outcomes are modeled by each random forest. The maximum value is one, which would indicate a perfect model, and the value can be negative if the models poorly simulate the observations. R^2 was calculated using the Scikit-learn random forest regressors score method. A random forest model that performs well should have a high R^2 score, and a low MAE. R^2 was primary used to validate that the MAE.

A best model, called the baseline model, was chosen from the brute force examples by sorting the models by their MAE. High accuracy models have low MAE values. A selection of top preforming models is presented in Table 7 and discussed in Section 6.1.

5.5. Scenarios for Sensitivity Analysis

To determine the accuracy of the baseline model, and to look at how small changes in hyperparameters effect random forest accuracy several scenarios for sensitivity analysis were performed. These can be broken down in to two categories: 1) feature changes and 2) hyperparameter changes. The feature changes were designed to look at how model results, and feature importance scores change when various features are removed from, the training feature set. The hyperparameter changes examine how small changes to the hyperparameters affect the model accuracy to better determine how the benefits of changing each hyperparameter. The scenarios for sensitivity are presented in Table 6.

Table 6: Scenarios for Sensitivity Analysis

Name	Type	Change
Thermokarst Initiation Model as Feature	feature change	Add Thermokarst Initiation Model (training labels) as input feature
Random Data as Feature	feature change	Add randomly generated data as an input feature
Remove Lat/Long	feature change	Remove northing and easting from input features
Remove Elevation, Slope and Aspect	feature change	Remove features with very low feature importance
Remove Seasonal Precipitations	feature change	Remove features that may be included as part of other features
Remove Other Low Impotence Features	feature change	Remove other features that may not be important
Remove High Impotence Features	feature change	Remove Highly important features to see how model importance is affected
Remove Other Low Impotence Features and Lat/Long	feature change	Remove Highly important features, and geolocation, to see how model importance is affected
Vary estimators	Hyperparameter change	Vary Estimators based on Baseline Random Forest
Vary Maximum Tree Depth	Hyperparameter change	Vary Tree depth around the value based on Baseline Random Forest
Vary Maximum decision features	Hyperparameter change	Vary number of features considered at each decision node based on Baseline Random Forest
Vary Maximum Leaf Nodes	Hyperparameter change	Vary maximum leaf nodes based on Baseline Random Forest
Vary minimum samples to split node	Hyperparameter change	Vary minimum samples to split node based on Baseline Random Forest
Vary minimum samples per leaf	Hyperparameter change	Vary minimum samples per leaf based on Baseline Random Forest
Vary Percent of training data used	Hyperparameter change	Vary Percent of training data used based on Baseline Random Forest

6. Results

6.1. The Baseline Random Forest

Choosing a baseline model from the many trained in the brute force period was done by sorting the models by their MAE. The model with the lowest MAE was selected, along with 10 other randomly picked models with low MAE to ensure selected models had a wide range of hyperparameter variation. These models were examined in more detail to determine which model was best to use as a baseline model for further testing.

While selecting the baseline random forest model it was discovered that the FDD data used for training the brute force models was mislabeled, making it off by a year. This created inaccurate results, which was especially apparent from the original feature analysis scores which did not equally score all climate features as expected, though it did not affect the MAE or R^2 scores. To correct the error, the 11 baseline candidate models were retrained with the corrected data. These retrained models had similar performance to the incorrect models, but with more equally scored all climate features. The similar performance these new models allowed them to remain baseline candidates.

The accuracy of all baseline options was examined over time, and all candidates had similar performance year to year with three options have similar slightly worse accuracy. (Figure 9). The baseline model could be chosen from the candidates based on the highest R^2, or the lowest MAE. Table 7 summarizes the accuracy of these models ordered and labeled according by MAE. All models scored greater than .95 for their R^2, and all had a MAE less than 2.0. Three options for a baseline model became apparent from these scores as indicated by the first 3 rows of Table 7. Two of these options had the same scores, and the lowest MAE (1.2866). The other option had the highest R^2 (0.9802). MAE was selected as the criteria to choose the model because the R^2 scores were similar for the all options. For the two remaining options all hyperparameters were the same except for max depth which was 60 or 100. The model with a max depth of 100 was selected as the baseline model. The baseline models hypermeters are summarized in Table 8.

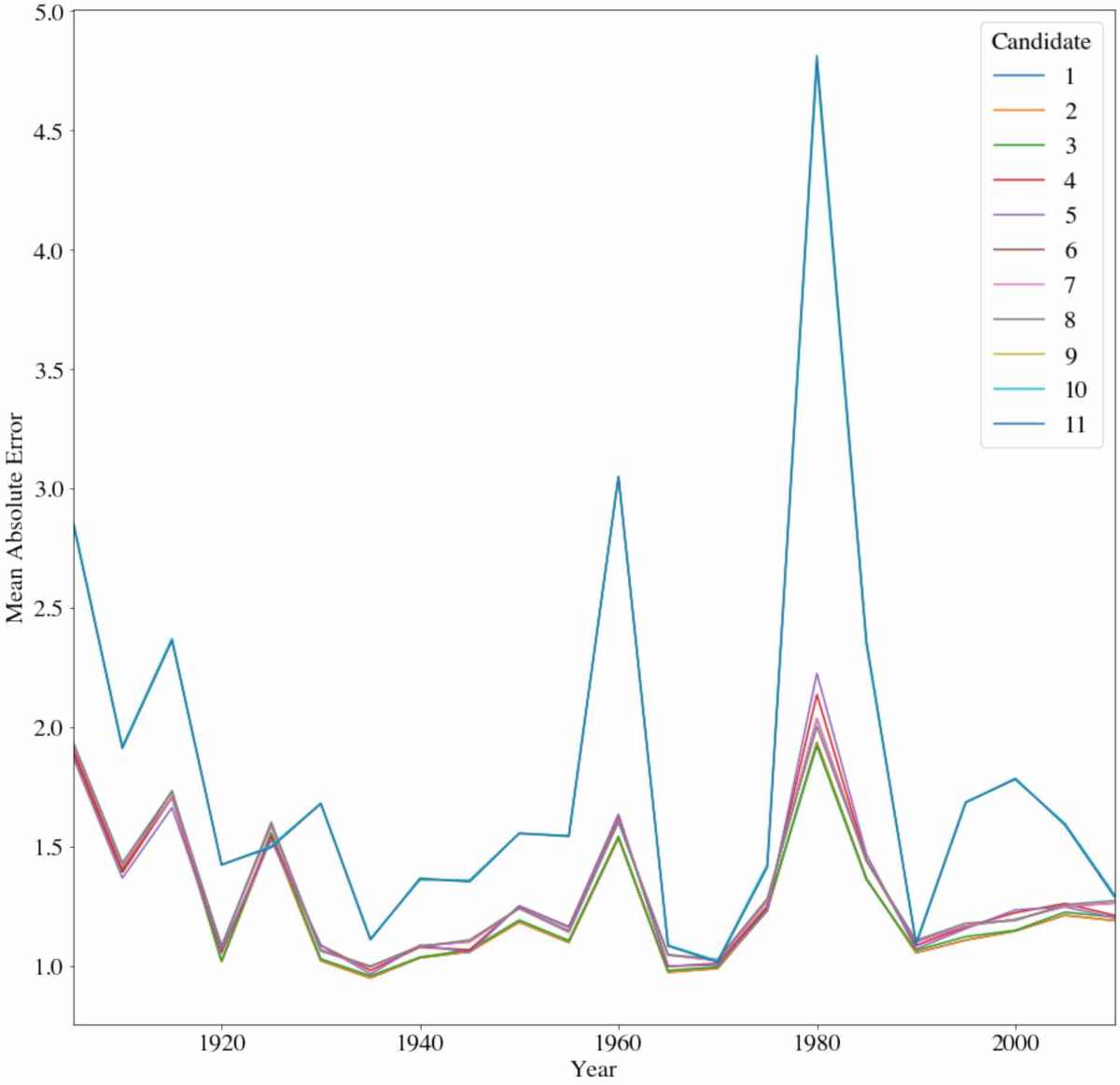

Figure 9: Comparison of candidate model accuracy for every 5 years. Each model predicts similarly for all years tested. Candidate numbers here correspond to Table 7. The table of values for this chart clarifies over lapping lines, and is in appendix C.

Table 7: Accuracy of baseline candidates sorted by MAE.

Candidate	Estimators	Maximum Tree Depth	Number Decision Features Considered	Maximum Leaf Nodes	Minimum Sample for leaf	Minimum Samples to split	Training data used	MAE	R^2
1	50	100	ALL (13)	50000	8	5	75%	**1.2866**	0.9801
2	50	60	ALL (13)	50000	8	5	75%	**1.2866**	0.9801
3	50	60	ALL (13)	50000	2	10	75%	1.2929	**0.9802**
4	50	60	ALL (13)	50000	4	5	25%	1.3393	0.9771
5	50	60	ALL (13)	50000	8	5	25%	1.3415	0.9759
6	10	60	ALL (13)	50000	8	2	50%	1.3418	0.9773
7	10	60	ALL (13)	50000	8	5	50%	1.3418	0.9773
8	10	100	ALL (13)	50000	4	10	50%	1.3437	0.9777
9	100	25	ALL (13)	50000	2	5	25%	1.8124	0.9499
10	100	25	ALL (13)	50000	2	2	25%	1.8127	0.9499
11	100	25	ALL (13)	50000	4	10	25%	1.8139	0.9493

Table 8: Baseline Hyperparameters

Hyperparameter	Value
Estimators	50
Maximum Tree Depth	100
Number Features Considered for decisions	All (13)
Maximum Leaf Node	50,000
Minimum Samples Needed to Split Node	5
Minimum Samples per Leaf	8
Training Data used	75%

Once the final baseline model was selected it was trained on the Seward Peninsula. The feature importance scores for both regions are shown in

Table 9. The numbers vary slightly for each region, but they have the same basic order with the exception of northing and next summer's TDD. For the ACP the MAE was 1.29 while it was 0.36 for the SP. R^2 was 0.980 for the ACP and 0.997 for the SP. These values indicate the model trained for the SP were more accurate with the same hyperparameters. These models were used as a baseline for all sensitivity analysis performed. Examples of the baseline model results, and the per pixel error are shown in Figure 10 and Figure 11.

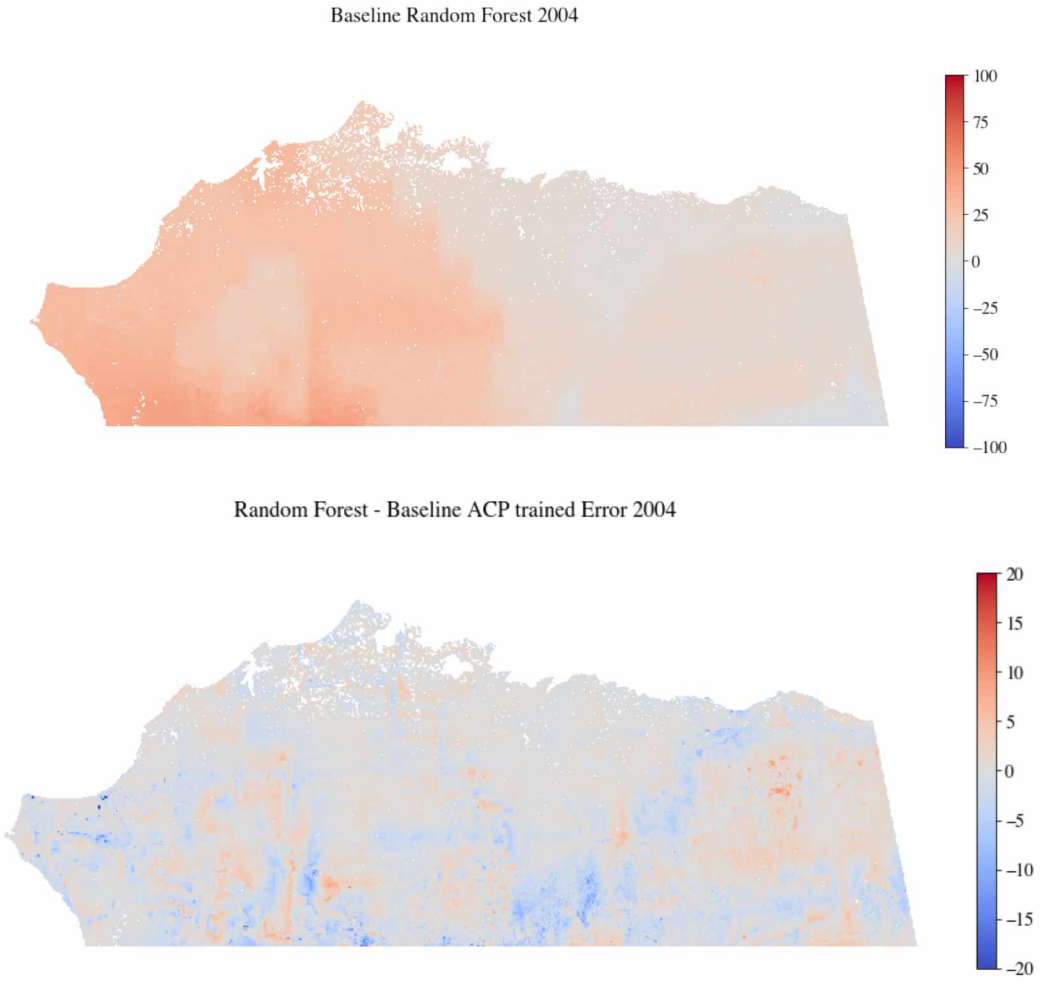

Figure 10: Baseline ACP results 2004, and the pixel error from the Original TKI model

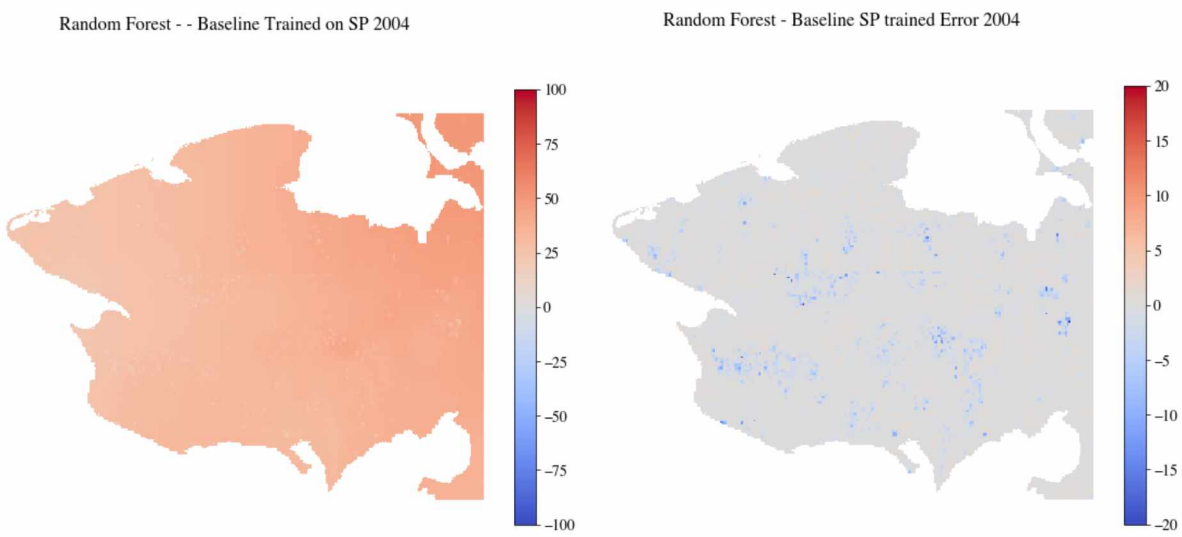

Figure 11: Baseline SP results 2004 and the pixel error from the Original TKI model

Table 9: Baseline feature importance scores

Feature	ACP Baseline RF model	SP Baseline RF model
Late Summer Precipitation	13.34	14.61
Next Summer Precipitation	12.75	13.12
FDD	11.70	11.49
Early Winter Precipitation	9.91	11.75
TDD	9.71	10.50
Winter Precipitation	9.58	11.20
Summer Precipitation	9.21	9.09
Northing	8.40	5.14
Next Summer TDD	8.13	7.98
Easting	7.24	5.11
Slope	0.02	0.02
Elevation	0.00	0.00
Aspect	0.00	0.00

6.2. Feature changes

The feature-based sensitivity analysis looks at how changes to the input features affects model accuracy. These scenarios were designed to look at random forest robustness, and how random forest models can be used to develop reduced order models. These scenarios are summarized in the first half of Table 6, and the resulting MAE and R^2 are presented in Table 10. The first two scenarios added input features at training, and primary looked at random forest model capabilities. The third scenario, removing northing and easting was intended to find if a location agnostic model trained on another was as accurate as a location-based model. The remaining scenarios looked at the robustness of random forest models as various features are removed from the training data.

Table 10: Summary of Feature Based Model Accuracy

	Seward Peninsula		**ACP**	
Model	**MAE**	**R^2**	**MAE**	**R^2**
Baseline	0.36	0.9973	1.29	0.98
Added Climate Priming TKI as input feature	0.00	1.00	0.00	1.00
Added Random Data as input feature	0.36	0.99	1.34	0.98
Removed Northing and Easting from input features	0.47	0.99	2.06	0.94
Removed Elevation, Slope, and Aspect from input features	0.36	0.99	1.30	0.98
Removed Partial Precipitations from input features	0.57	0.99	1.67	0.96
Removed Least Important Features from input features	0.57	0.99	2.62	0.91
Removed Most Important Features from input features	0.73	0.99	1.97	0.95
Removed Most Important Features, and Northing, Easting from input features	1.65	0.96	3.65	0.83

All of these feature bases scenarios examine adding or removing features and how the random forest models created responded to those changes. Removing features generally created poorer models with accuracy decreasing as more features were removed. Adding the climate priming TKI to the input features created models where the decision process largely relied on the TKI feature with MAE and R^2scores produced near 0 and 1 respectively with the TKI model having 100% feature importance , as is expected if RF models can make decisions on important features. Conversely, when adding random data to the input features the random data was ignored (accounting for 0% of feature importance), and the MAE and R^2scores changed little. Selected tables showing how feature importances changed in these scenarios are in Appendix C.

The third scenario, removing Northing and Easting was intended to show if a location agnostic version of a random forest model could be created for determining TKI values. Applying these new models to the region they were trained on produced results very similar to the baseline model(Table 10) and low error(Figure 12). On the other hand, when these models were applied to the other region (ACP model with SP training data, and vice versa) the results contained larger errors(Figure 13) making them not useful outside the region they were trained.

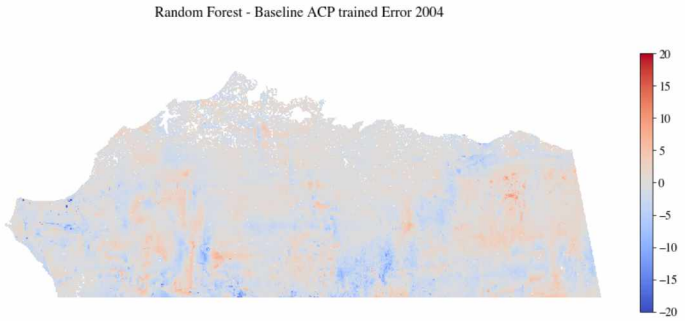

Figure 12: ACP random forest model (no northing/easting) on ACP

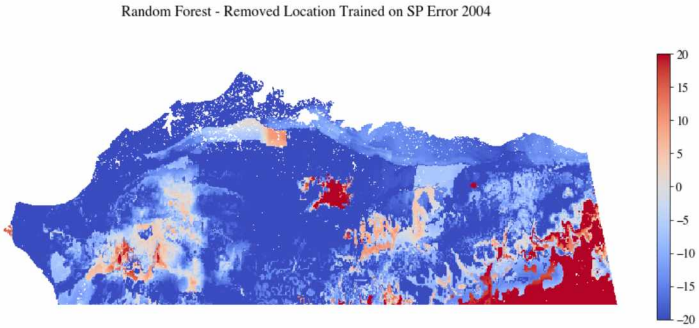

Figure 13: Seward Peninsula random forest model (no northing/easting) on ACP

6.3. Hyperparameter changes

Adjusting Hyperparameter values and retrain a random forest model with the same data creates results that vary based on the change. Sensitivity analysis was performed to determine the best settings for hyperparameters and determine how adjusting them around the baseline value changes RF accuracy. Settings tested in scenarios and resulting best values for modeling TKI with random forests are presented in Table 11. For each scenario, only the hyperparameter being tested was changed. Charts depicting how the MAE and R^2 scores changed in each scenario are in Appendix D.

One key finding of this analysis is that one parameter should be selected to control tree size. There are two options that can control the size of each tree: maximum tree depth, and maximum leaf nodes. Maximum depth allows tress to grow to maximum of 2^{depth} leaf nodes while maximum leaf nodes allows growth to a maximum of it's value. Whichever of these hyperparameters is smaller will be the determining factor for tree size. For the sensitivity tests performed as part of this report maximum leaf nodes was almost always the limiting factor, as maximum depth, for the baseline RF model, was set incredibly large at 2^{100}. Either option may be chosen, but maximum leaf nodes was selected here.

Other findings include that using more estimators, and choosing a decision feature from almost all of the possible options at each decision node create better results . Better performance with more estimators is expected as combing the results of many trees should produce results that more closely replicate the problem being modeled (Breimen 2001). Considering most but not all of the decision features at each decision node produced the best results in this analysis. As features are considered the accuracy of models increased until around ¾ of the options were possible, after that increasing the value slightly decreased model accuracy (Figure 14).

Table 11: Summary of Hyperparameter Sensitivity Analysis

Hyperparameter	Values tested	Best Value
Estimators	1, 4, 12, 25, 50, 60, 75, 100	50
Maximum Tree Depth	1-20, 30, 40,50,60,70,80,90,100	Leave as default which allows best possible max tree depth given other parameters.
Number Features Considered for decisions	1, 3, 4, 5, 6, 8, 10, 12, 13	11

Maximum Leaf Nodes	12500,25000,37500,50000, 62500, 75000, 10000	50000
Minimum Samples Needed to Split Node	2, 5, 7, 10, 15, 25	Any
Minimum Samples per Leaf	1, 4, 8, 12, 16	Any
Training Data Percent	25%, 50%, 75%	75%

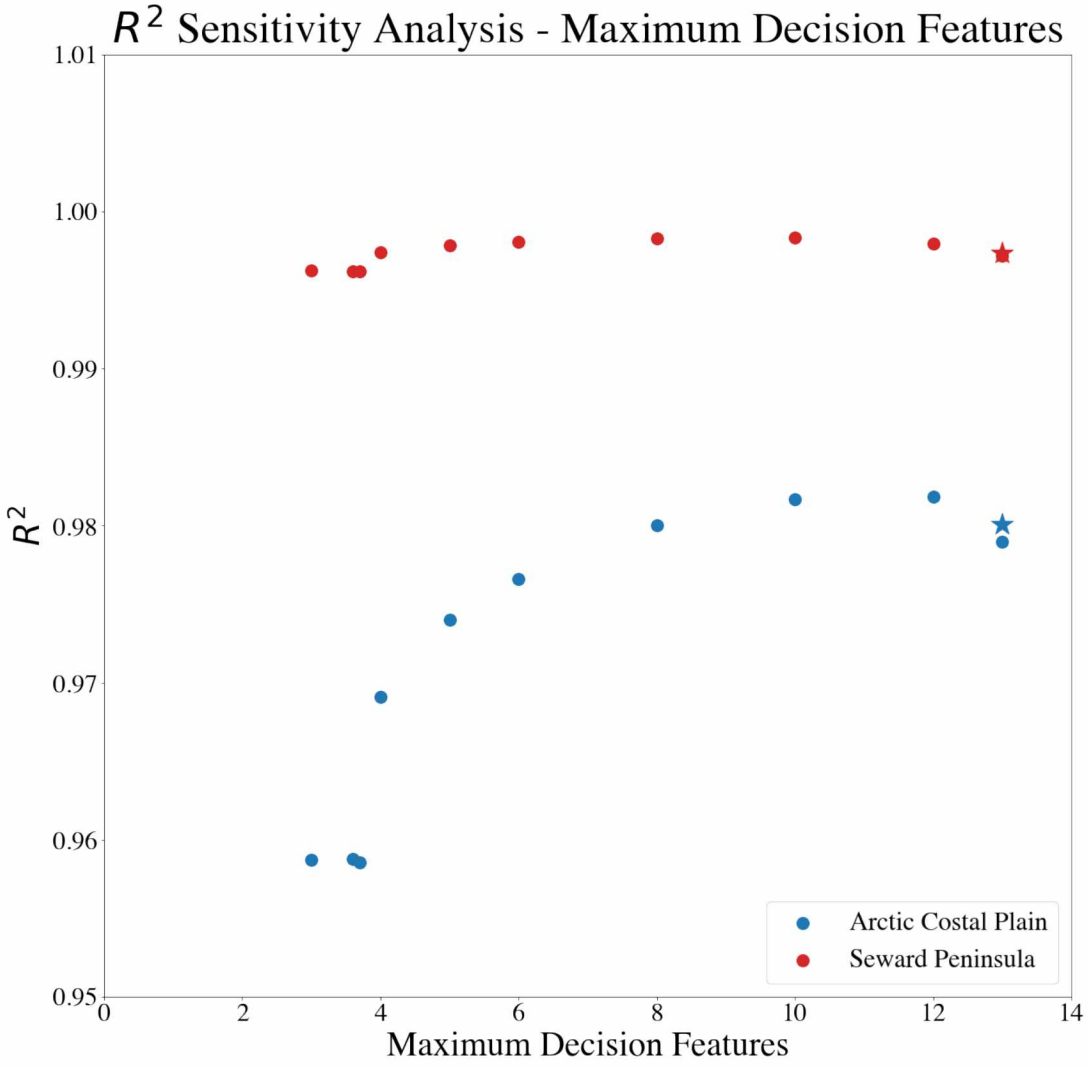

Figure 14: R^2 for maximum fecision feature sensitivity analysis showing decreases in accuracy when hyper parameter is set greater than 10

6.4. A Final Model

Using the insight gained in the feature based, and hyperparameter sensitivity analyses a final set of models was developed. These models were trained for each region separately. As indicated by the feature-based analysis was used to train the models. The hyperparameters used were very similar to the base line model. Key differences include only using maximum leaf nodes to control tree size and using 11 as the number of decision features considered. Max leaf nodes was selected as it was more apparent where the increasing value became less useful than adjusting Max depth. A value of 75,000 was used for this parameter. 11 was selected for the decision feature consideration because it was between the two values that produced the best results during the sensitivity analysis. The hyperparameters used in the final model are summarized in Table 12.

Table 12: Summary of Final Hyperparameters

Hyperparameter	Baseline	Final
Estimators	50	50
Maximum Tree Depth	100	None
Number Features Considered for decisions	Auto (13)	11
Maximum Leaf Nodes	50,000	75,000
Minimum Samples Needed to Split Node	5	5
Minimum Samples per Leaf	8	8
Training Data Percent	75%	75%

The final model yielded better results than the baseline mode for both regions. There was a significant decrease ($\Delta - 0.1808$ for SP, $\Delta - 0.3244$ for the ACP) in the MAE showing that predictions were closer to the expected value. The increase in the R^2 score was less apparent, but those values are already very high. Table 13 summarize the final accuracy results.

Table 13: Final Accuracy Results

	SP		ACP	
Measure	Baseline	Final	Baseline	Final
MAE	0.3598	0.1790	1.2866	0.9662
R^2	0.9973	0.9983	0.9801	0.9885

7. Discussion

The feature-based sensitivity analysis yielded three important results. The first is that random forest models are able to successfully find important features and use them to inform their decision process. The second is that models are not a useful tool for elimination input features if all features are equally important to the decision process. Finally, this analysis showed that the models must be retrained for each region they are required to be run on.

Random forest models are a robust tool for finding relationships between many possible inputs and a given output. This is especially supported by the sensitivity analysis where the original TKI model, and random data were added. When the TKI model was added the RF process the model used it as the main driver of the decision process. While when the random data was added it was ignored. The robustness of random forest models is supported to a lesser extent by the fact that the random forest models assigned equal weight to features used to train the original TKI model.

One of the purposes of using random forest models to replicate the thermokarst imitation model was to determine if they are a useful tool to develop reduced order models. From the results of the feature-based analysis it was not immediately apparent that this is possible. All of the features from the original model were indicated as being equally important to the random forest decision process, as indicated by the feature importance values. Additionally, removing any of these features from the training process reduced model accuracy. Conversely removing features that were not important (slope, elevation, and aspect) did not affect model accuracy. This shows that using random forests to develop a reduced order model may be possible if any of the inputs to the original model were shown to be considerably less important than the other inputs. Performing these analyses with other models would need to be completed to determine if this is truly the case.

As well as being less complex, it is also desired that reduced order models be more efficient in time. While useful timing data for comparing was not collected, all random forest models trained took less than 8 hours to train with 1.77 hours for the final ACP model. This may not seem efficient given that the original TKI model takes a few minutes to generate predictions, but the prediction time for the random forest models took 45 seconds for the final ACP model. The benefit would come from applying the random forest model to different data sets (for example using projected data to create future forecasts) though as other results have shown this is limited to single regions for this project.

An attempt to develop a location agnostic model was attempted during the feature-based sensitivity analysis. When northing and easting were removed model, accuracy was not significantly reduced, but this only applied to the region for which the model was trained. These models did not perform well for the other sample region. This shows that random forest models should be limited to the regions for which they are trained. This could also lead to limitations to this RF process if continued climate change causes changes to temperature or precipitation that are outside of the ranges of values seen in the training data.

The hyperparameter sensitivity analysis showed that the most important hyperparameters are indeed the number of estimators, and number of decision features. The analysis showed that increasing the values for these features increased model accuracy in the quickest fashion. The results also indicated that using only one hyperparameter to control tree size is advisable either choose maximum leaf nodes or maximum tree depth. Finally changing training data percentages increases the training time in a linear way, but for this problem accuracy did not increase in an extreme way as more data was added.

One thing that using a random forest model may be useful for is determining if assumptions about inputs made in an existing model are accurate. A related use would be using a random forest model of a complex model, to determine if running the full model is worthwhile. For the TKI process all of the original inputs were indicated as being equally important during the analysis. This was expected as the original TKI climate priming model weighed all inputs equally. If this had not been the case it would indicate that the original model would need to be reworked. Again, preforming this type of analysis on other, perhaps more complex models, would verify the usefulness of this process.

8. Future Work

Designing a testing framework that give useful timing data, for training and evaluation, would be a good future project that would give another point of comparison for random forest models. Random forest models can accurately simulate other models, but from the data gathered here it is difficult to determine how long training will take when a hyperparameter is changed. Better timing data would allow for more informed decisions when choosing models if time efficiency is important. This could be important for more complex models than the TKI model examined here.

Further improving the underlying TKI model and retraining the random forest models can give new data to show the validity of random forest models. The original TKI model is currently uncalibrated and lacks meaningful verification from real world

observations. These observations are hard to come by due to lack in existing data for the regions being compared. When the original model's results are verified more concrete statements can be made about random forest accuracy. Additionally, using the random forest models developed here to predict future TKI values would help to determine the robustness of the random forest process. Values predicted from projected data could be compared to values with similar input features to determine validity.

Random forest models are a good tool for simulating the TKI process examined here, but they could also be a useful tool for simulating other parts of the full ATM. Many parts of the ATM are more complex than the TKI component, and a random forest model that could be pretrained and integrated into the ATM could save time when running the full ATM. They may also be useful in determining which inputs to the ATM, or other complex model, are truly important through the feature importance scores. This may aid in the development of ROMs. Random forests could also be used to create quick initial results from new data sets to see if they may be useful if run in the full model.

Another improvement to these random forest models may come from applying them to smaller regions. The SP models performed better than the ACP model in these tests. To determine if smaller regions are truly effective the ACP could be split into smaller regions, and a random forest could be trained for each. The ACP and SP data could be combined. Based on this hypothesis the smaller regions should have better accuracy while the larger should have poorer performance.

Random Forest models can have similar performance to other machine learning methods like SVMs and ANNs, but best accuracy is application dependent (Cutler, et al. 2007) (Were, et al. 2015) (Pal 2007) (Wang, et al. 2016). Applying those methods this problem could determine if RF was the best choice for this project. Random forests do have two advantages the other methods do not. They are simpler to use requiring less parameters to be set (Pal 2007) (Wang, et al. 2016), and they provide a variable importance measure that the other methods do not (Cutler, et al. 2007) .

9. Conclusion

Random forest models were able to simulate the results of an existing model as shown by their application to TKI modeling. It was also shown that random forest models are able to find important features and ignore unimportant ones. Sensitivity to changes in hyperparameters was examined. From this, it was determined that the number of estimators, number of features considered at each node are important to the

process which is in agreement with future work. Finally, the interaction between hyperparameters was examined which showed that only one of these should be set.

Looking at random forests showed their value for verifying assumptions in existing models. They can also be used in and for the development of reduced order models. The feature importance scores can be used to demine which inputs may be eliminated, and for very complex models they may be used as a ROM, as they only need to be trained once, and prediction times are generally quick. Though this project only looked a applying random forests to the TKI part of the ATM they would also be useful in other portions of the ATM.

10. Acknowledgments

The research presented in this work was funded by The Alaska Climate Adaptation Science Center and the Department of Energy Office of Biological and Environmental Research Next Generation Ecosystem Experiment (NGEE-Arctic) project.

11. Code and Additional Examples

Code and additional examples can be fount on the GitHub repository for this project at https://github.com/rwspicer/masters-project. Additionally the ATM can be found at https://github.com/ua-snap/arctic_thermokarst_model.

12. References

Antoulas, A.C. 2004. "Approximation of Large-Scale Dynamical Systems: An Overview." *IFAC Proceedings Volumes* 37 (11): 19-28.

Breimen, Leo. 2001. "Random Forests." *Machine Learning* 45: 5-32.

n.d. *Conda.* Accessed 02 05, 2019. https://conda.io/projects/conda/en/latest/.

Cutler, Richard D., Thomas C. Edwards, Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler. 2007. "Random Forests for Classifaction in Ecology." *Ecology* 2783-2792.

Davis, Neil. 2001. *Permafrost.* Fairbanks, Alaska: University of Alaska Press.

DGGS. n.d. *DGGS elevation portal.* Accessed 11 02, 2019. https://elevation.alaska.gov/.

Farquharson, L.M., D.H. Mann, Grosse G., Jones B.M., and Romanovsky V.E. 2016. "Spatial distobution of thermokarst terrain in Arctic Alaska." *Geomprphology* 273: 116-133.

Gandodamage, Chandana, Rowland Joel C., Susan S. Hubbard, Steven P. Brumby, Anna K. Liljedahl, Harukdo Wainwright, Cathy J. Wilson, et al. 2014.

"Extapolating active layer thickness measurements across Arctic poolygoonal terrain using LiDAR and NDVI data sets." *Water Resurces Research* 50: 6339-6357.

Hastie, Trevor, Robert Tibshirani, and Friedman Jerome. 2017. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction Second Edition.* Springer.

Herman, Gregorgy R., and Russ J. Schumacher. 2018. "Money Doesn't Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation WIth Randm Forests." *America Meteorlogical Society* .

Jorgenson, M. T., Kenevskiy M., Shur Y., Moskalenko N., Brown D. N. R., Wickland K., R. Striegl, and Koch J. 2015. "Role of ground ice dynamics and eclgical feedbacks in recent ice wedge degradtion and stabilization." *Joournal of Geophysical Research: Earth Surface* 2280-2297.

Jorgenson, M. Torre, Yuri L. Shur, and Erik R. Pullman. 2006. "Abrupt Increase in permafrost degradion in Arctic Alaska." *Geophysical Research Letters* 33.

Kanevskiy, Mikhail, Yuri Shur, Torre Jorgensoon, Dana R.N. Brown, Nataliya Moskalenkom, Jerry Brown, Dnald A. Walker, Martha K. Raynolds, and Marcel Buchhorn. 2017. "Defradation and Stabilization of Ice Wedges Implications for Assessing Risk of Thermokarst in Northern Alaska." *Geomorphology* 297: 20-42.

Lara, Mark J., McGuire A. David, Eugenie Euskirchen, Tweedie Craig E., and Kenneth M., Skurikhin Alexei N., Romanivsky Valdimir E., Grisse Guido, Bolton, W. Robert, Genet, Helene Hinkel. 2014. "Polygonal Tundra Geomrpholoogical Change in Respnse to Warming Alters Future CO2 and CH4 Flux on The Barrow Peninsula." *Golbal Change Biology.*

Liljedahl, A.K., J. Boije, R.P. Dannen, Fedorov A.N., G.V. Frost, G. Grosse, L.D. Hinzman, et al. 2016. "Pan-Arctic ice-wedge degradtion in warming permafrost and its influence on tundra hydroloogy." *Nature Geoscience* 9.

McGovern, Amy, Kimberly L. Elmore, David Jonh II Gagne, Sue Ellen Haupt, Christopher D. Karstens, Ryan Lagerquist, Travis Smith, and John K. Williams. 2017. "Using Artifical Intelligence to Improve Real-Time Decision-Making for High-Impact Weather." *American Meteorological Society.*

Pal, M. 2007. "Random forest Classifier for remote sensing classification." *Internatinal Journal f Remote Sensing* 26: 217-222.

Pedregosa, F., G Varoquaux, A Gramfrt, V Michel, B Thirin, O Grisel, M Blondel, et al. 2011. "Scikit-learn: Machine Learning in Pythn." *Journal of Machine Learning Research* 2825-2830.

Pollard, W. 2018. "Chapter 15 - Periglacial Processes in Glacial Environments." In *Past Glacial Environments*, edited by John Menzies and Jaap J.M. van der Meer, 537-564. Elsevier.

Raynolds, Martha K., Donald A. Walker, Kenneth J. Ambrosius, Brown Jerry, Kayer Everett, Kanevskiy Mikhail, Gary P. Koofinas, Vladimr E. Romanovsky, Yuri Shur, and Patrick J. Webber. 2014. "Cumulative geoecological effects of 62 years

of infrastructure and climate change in ice-rich permafrost landscapes, Prudoe Bay Oilfeild, Alaska." *Global Change Biolgy* 1211-1224.

Robert, Bolton W., Hélène Genet, Mark Lara, Valdimir Romanovsky, and William Riley. 2018. "Identifying Regions Susceptible to Thermokarst Initiation on the Alaskan Arctic Coastal Plain for the Alaska Thermokarst Model."

Rowley, Taylor, John R. Giardino, Raquel Granados-Aguilar, and John D. Vitek. 2015. "Chapter 13 - Periglacial Processes and Landforms in the Critical Zone." Edited by John R. Giardino and Chris Houser. *Developments in Earth Surface Processes* (Elsevier) 19: 397-447.

Scenarios Network for Alaska and Arctic Planing. 2019. "Historical Monthly Temperature - 1 km CRU TS." University of Alaska. Accessed 11 04, 2019. http://ckan.snap.uaf.edu/dataset/historical-monthly-temperature-1-km-cru-ts.

Scenarios Network for Alaska and Arctic Planning. 2019. "Historical Monthly Precipitation - 1 km CRU TS." University of Alaska. Accessed 11 04, 2019. http://ckan.snap.uaf.edu/dataset/historical-monthly-precipitation-1-km-cru-ts.

scikit-learn-developers. n.d. *3.2.4.3.2. sklearn.ensemble.RandomForestRegressor - scikit-learn 0.21.3 documentation.* Accessed 02 05, 2020. https://scikit-learn.org/0.21/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor.

spatialreference.org. 2006. *Spatial Reference epsg projection 3338.* Accessed 02 12, 2020. https://spatialreference.org/ref/epsg/nad83-alaska-albers/.

Wang, Li'ai, Xudong Zhou, Xinkai Zhu, Zhadodi Dong, and Wenshan Guo. 2016. "Estimation of biomass in wheat using randm forest regressin algorithm and remotesensing data." *The Crop Journal* 4: 212-291.

Were, Kennedy, Dieu Tien Bui, Oystein B. Dick, and Bal Ram Singh. 2015. "A comparitive assesment of support vector regression, artifical neural networks, and random forests for predicting and mapping soil orgianc carbon stoocks across Afromontane landscape." *Ecooloogical Indicators* 394-493.

Appendix A: Example Precipitation Maps

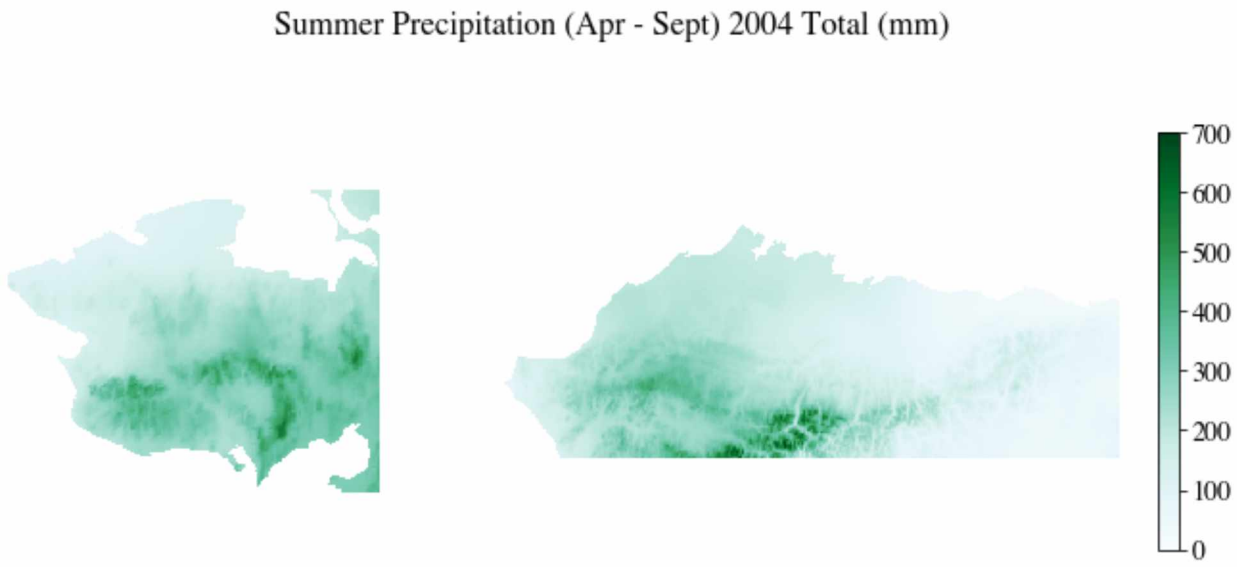

Figure 15: Summer Precipitation for 2004 (Seward Peninsula Left, ACP Right

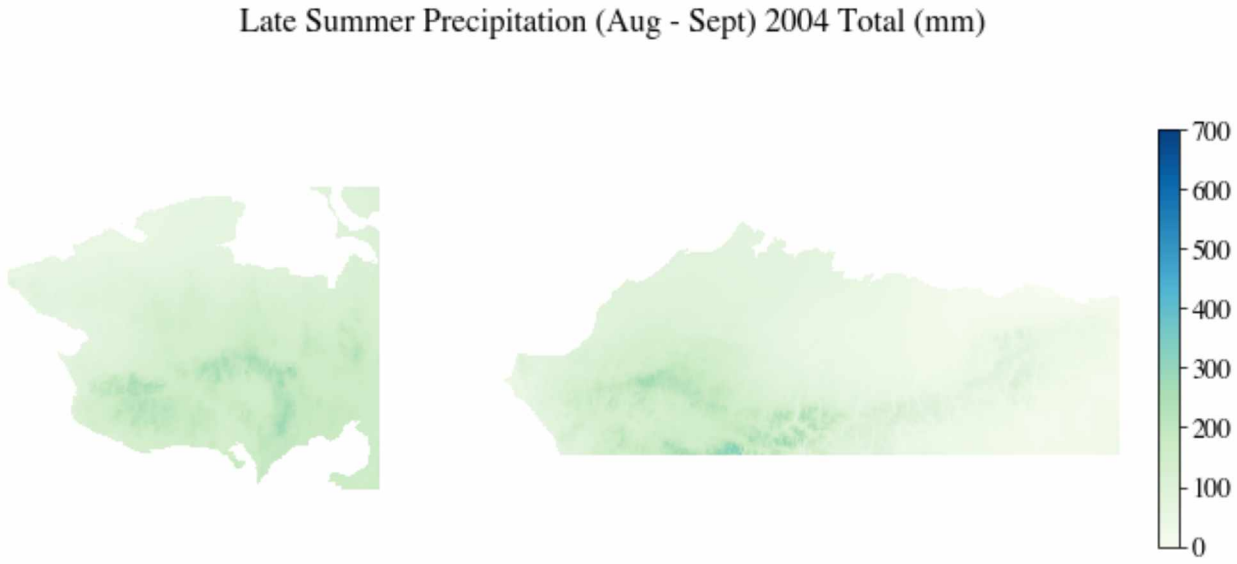

Figure 16: Late Summer Precipitation for 2004 (Seward Peninsula Left, ACP Right)

Figure 17: Early Winter Precipitation for 2004-2005 (Seward Peninsula Left, ACP Right)

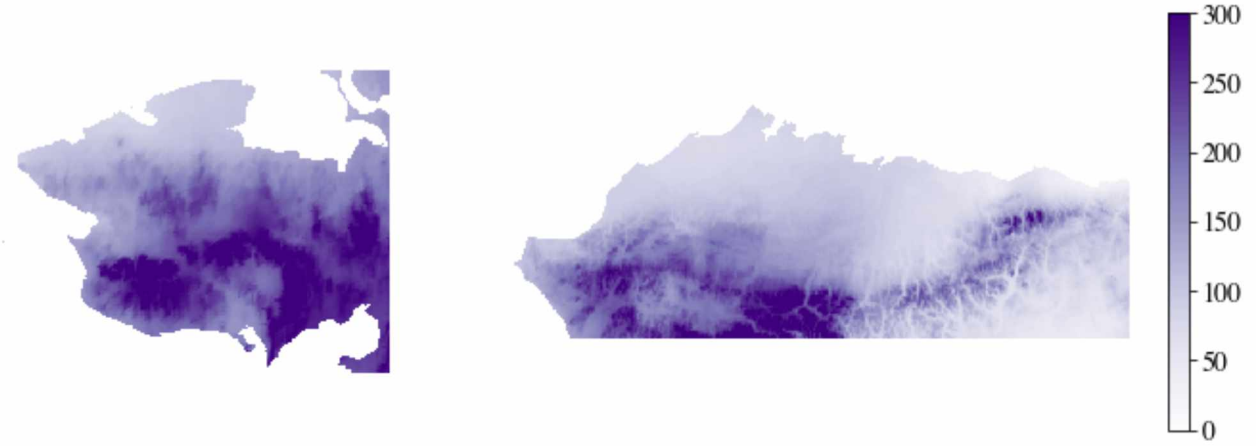

Figure 18:Total Winter Precipitation for 2004 - 2005 (Seward Peninsula Left, ACP Right)

Appendix B: Climate Priming Model TKI predictions

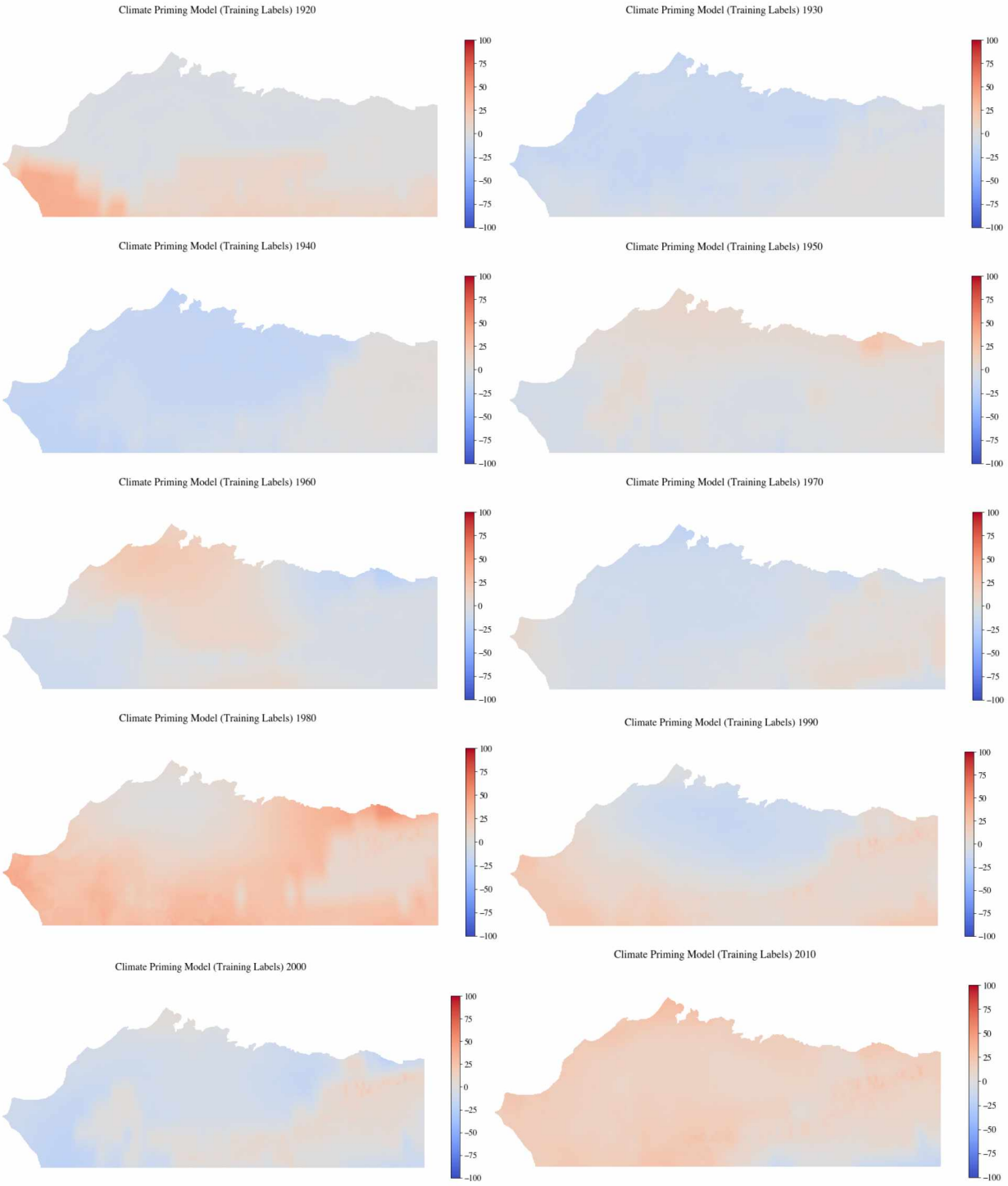

Figure 19: ACP TKI examples

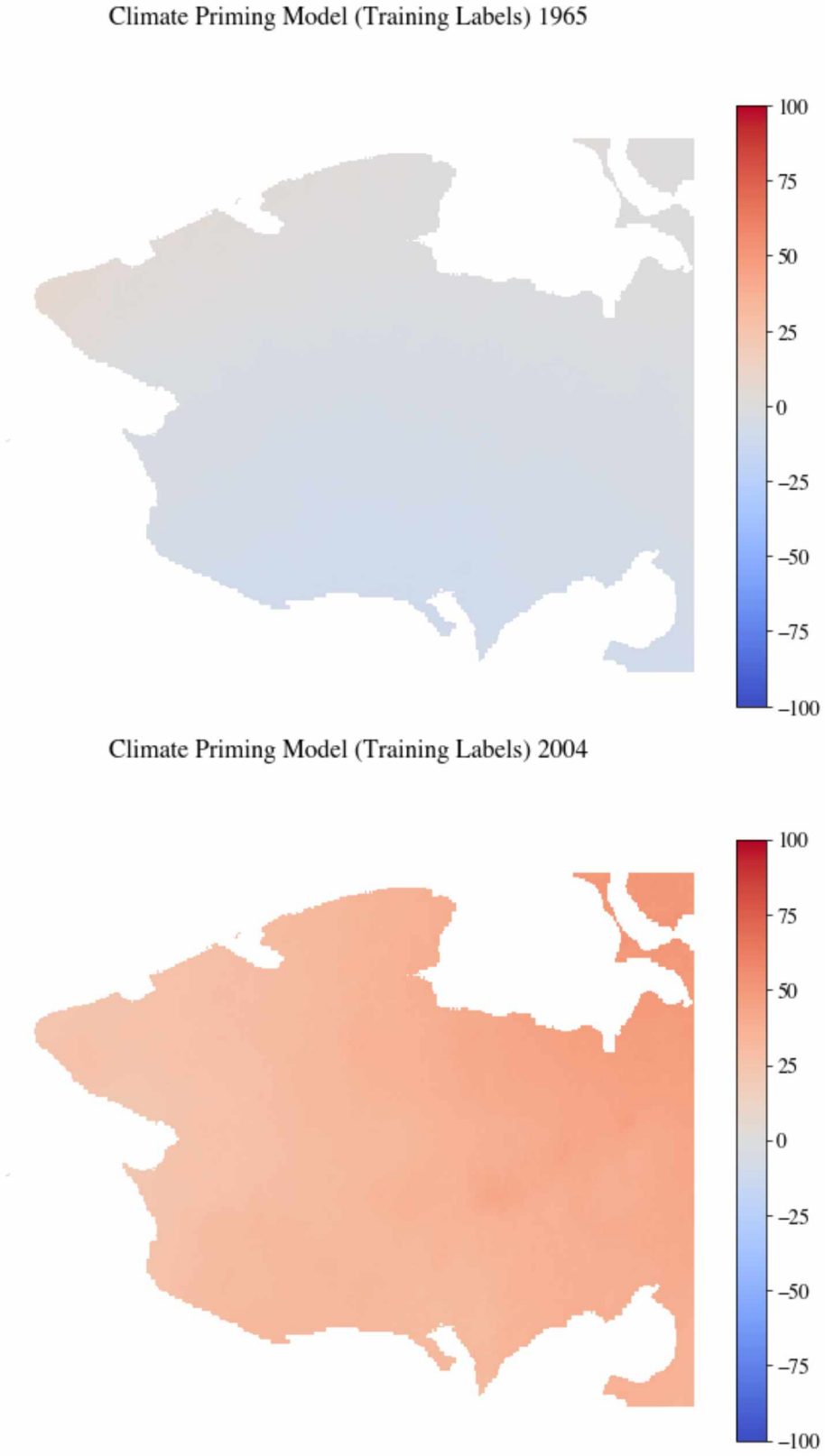

Figure 20: Seward Peninisula TKI for 1965 and 2004

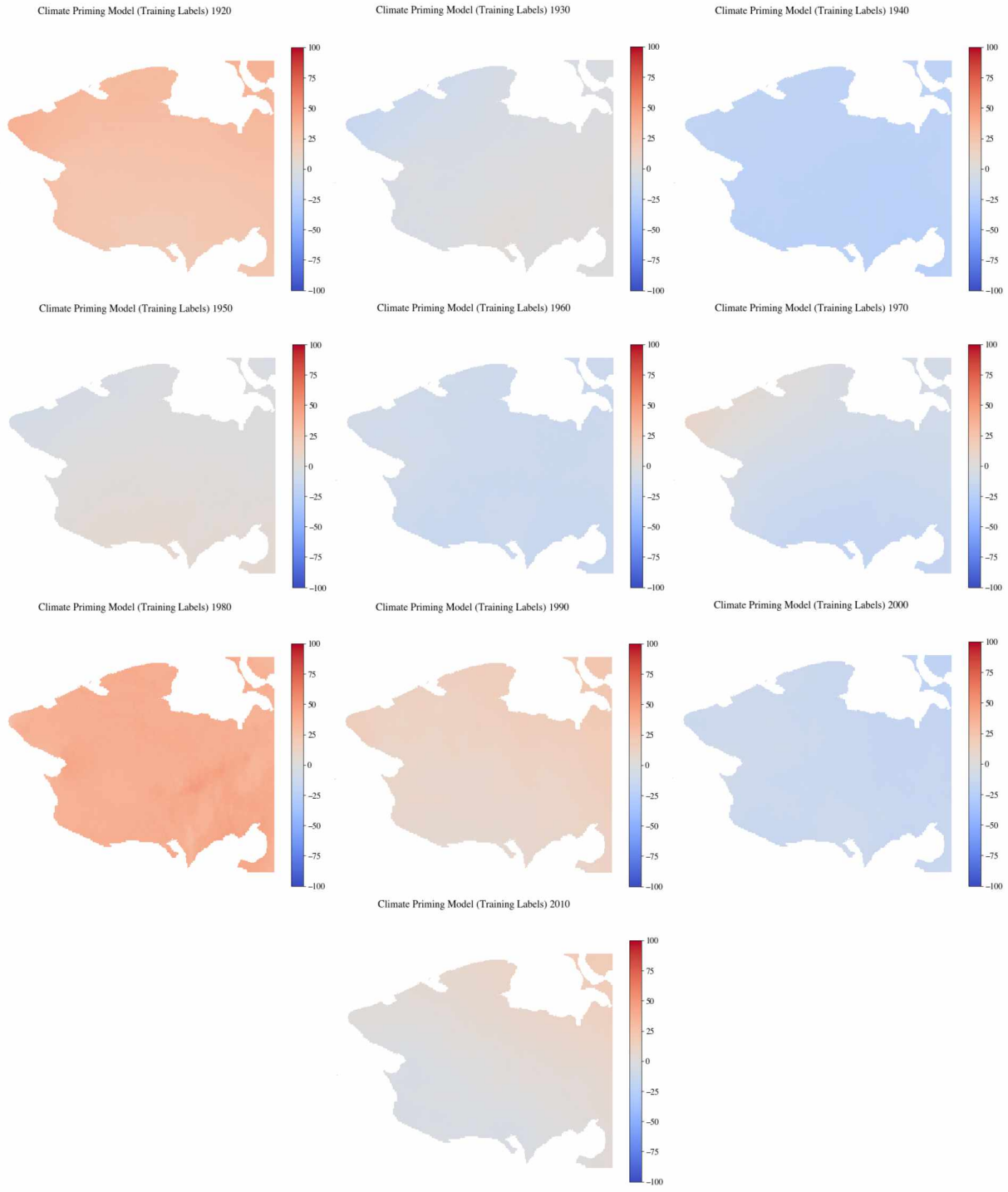

Figure 21: Seward Peninsula TKI examples

Appendix C: Additional Charts

Table 14: Data for Figure 9

	Candidates										
Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1905	1.888223	1.888223	1.905425	1.907239	1.877817	1.927964	1.927964	1.940959	2.867532	2.867263	2.862704
1910	1.393067	1.393067	1.406765	1.392940	1.367255	1.416609	1.416609	1.431508	1.917603	1.917964	1.910368
1915	1.707448	1.707448	1.732675	1.707389	1.662293	1.706748	1.706748	1.729944	2.370224	2.370789	2.362147
1920	1.017236	1.017236	1.020779	1.054837	1.065116	1.082016	1.082016	1.080586	1.421171	1.422320	1.424070
1925	1.539721	1.539721	1.557143	1.541329	1.528591	1.586656	1.586656	1.601436	1.500603	1.501046	1.495117
1930	1.020585	1.020585	1.028380	1.085329	1.085863	1.061630	1.061630	1.067020	1.680309	1.680865	1.678642
1935	0.948709	0.948709	0.957450	0.980221	0.965275	0.995467	0.995467	0.998039	1.112064	1.112529	1.108064
1940	1.032876	1.032876	1.035966	1.077688	1.086694	1.082972	1.082972	1.080501	1.359475	1.360498	1.366856
1945	1.058693	1.058693	1.065811	1.066423	1.055871	1.100259	1.100259	1.107853	1.358207	1.358293	1.351408
1950	1.182352	1.182352	1.191080	1.250076	1.250466	1.240312	1.240312	1.241955	1.554042	1.554343	1.555086
1955	1.097479	1.097479	1.105379	1.163697	1.160729	1.140497	1.140497	1.147368	1.544358	1.544717	1.541924
1960	1.534483	1.534483	1.542259	1.627509	1.635278	1.606918	1.606918	1.603288	3.049092	3.048864	3.050536
1965	0.972629	0.972629	0.980007	0.997034	0.998618	1.044811	1.044811	1.045980	1.084881	1.085390	1.082990
1970	0.988505	0.988505	0.994807	1.009420	1.003502	1.022426	1.022426	1.027071	1.020008	1.020111	1.014594
1975	1.232722	1.232722	1.244465	1.248790	1.228984	1.268774	1.268774	1.281650	1.422045	1.422230	1.412460
1980	1.935587	1.935587	1.921438	2.135983	2.225187	2.035310	2.035310	2.002289	4.779295	4.779186	4.813429
1985	1.365122	1.365122	1.359836	1.443615	1.465897	1.439667	1.439667	1.438489	2.346966	2.347087	2.357033
1990	1.054167	1.054167	1.064509	1.084649	1.070357	1.099927	1.099927	1.106404	1.096359	1.096766	1.089475
1995	1.107227	1.107227	1.121799	1.162570	1.155070	1.167069	1.167069	1.178222	1.685784	1.685880	1.682968
2000	1.145200	1.145200	1.148371	1.222172	1.233862	1.193945	1.193945	1.189282	1.780549	1.780174	1.784821
2005	1.211572	1.211572	1.224400	1.260649	1.246442	1.247881	1.247881	1.256056	1.594900	1.595534	1.589498
2010	1.189141	1.189141	1.203924	1.210605	1.201037	1.262280	1.262280	1.272627	1.296097	1.296826	1.288325

Table 15: Feature Importance Changes When Adding Climate Priming TKI to Training Features

Feature	ACP Baseline RF model	ACP Adding TKI RF Model	SP Baseline RF model	SP Adding TKI RF Model
Summer Precipitation	9.21	0.00	9.09	0.00
Late Summer Precipitation	13.34	0.00	14.61	0.00
Early Winter Precipitation	9.91	0.00	11.75	0.00
Winter Precipitation	9.58	0.00	11.20	0.00
Next Summer Precipitation	12.75	0.00	13.12	0.00
TDD	9.71	0.00	10.50	0.00
FDD	11.70	0.00	11.49	0.00
Next Summer TDD	8.13	0.00	7.98	0.00
Northing	8.40	0.00	5.14	0.00
Easting	7.24	0.00	5.11	0.00
Elevation	0.00	0.00	0.00	0.00
Slope	0.02	0.00	0.02	0.00
Aspect	0.00	0.00	0.00	0.00
TKI model	N/A	100.00	N/A	100.00

Table 16: Feature Importance Changes When Adding Random Data to Training Features

Feature	ACP Baseline RF model	ACP Adding Random Data	SP Baseline RF model	SP Adding Random Data
Summer Precipitation	9.21	9.19	9.09	9.24
Late Summer Precipitation	13.34	13.19	14.61	14.70
Early Winter Precipitation	9.91	9.92	11.75	11.80
Winter Precipitation	9.58	9.69	11.20	11.04
Next Summer Precipitation	12.75	12.80	13.12	13.29
TDD	9.71	9.46	10.50	10.41
FDD	11.70	11.72	11.49	11.42
Next Summer TDD	8.13	8.33	7.98	7.97
Northing	8.40	8.35	5.14	5.20
Easting	7.24	7.32	5.11	4.89
Elevation	0.00	0.00	0.00	0.00
Slope	0.02	0.02	0.02	0.02
Aspect	0.00	0.00	0.00	0.00
Random Data	N/A	0.00	N/A	0.00

Table 17: Feature Importance Changes When Removing Slope, Elevation, and Aspect From Training Features

Feature	ACP Baseline RF model	ACP Remove Slope, Elevation, and Aspect	SP Baseline RF model	SP Remove Slope, Elevation, and Aspect
Summer Precipitation	9.21	14.16	9.09	11.62
Late Summer Precipitation	13.34	12.61	14.61	17.20
Early Winter Precipitation	9.91	14.19	11.75	12.52
Winter Precipitation	9.58	10.89	11.20	12.56
Next Summer Precipitation	12.75	15.34	13.12	14.05
TDD	9.71	10.23	10.50	11.11
FDD	11.70	12.11	11.49	12.06
Next Summer TDD	8.13	9.42	7.98	8.72
Northing	8.40	N/A	5.14	N/A
Easting	7.24	N/A	5.11	N/A
Elevation	0.00	0.00	0.00	0.01
Slope	0.02	1.04	0.02	0.15
Aspect	0.00	0.00	0.00	0.01

Table 18: Feature Importance Changes When Removing Partial Season Precipitation from Training Features

Feature	ACP Baseline RF model	ACP Remove TKI RF Model	SP Baseline RF model	SP Remove TKI RF Model
Summer Precipitation	9.21	16.77	9.09	14.86
Late Summer Precipitation	13.34	N/A	14.61	N/A
Early Winter Precipitation	9.91	N/A	11.75	N/A
Winter Precipitation	9.58	11.87	11.20	15.00
Next Summer Precipitation	12.75	16.54	13.12	15.86
TDD	9.71	10.76	10.50	13.77
FDD	11.70	12.25	11.49	15.27
Next Summer TDD	8.13	12.36	7.98	11.15
Northing	8.40	7.85	5.14	6.87
Easting	7.24	11.60	5.11	7.01
Elevation	0.00	0.00	0.00	0.02
Slope	0.02	0.07	0.02	0.19
Aspect	0.00	0.00	0.00	0.01

Appendix D: hyperparameter sensitivity analysis charts

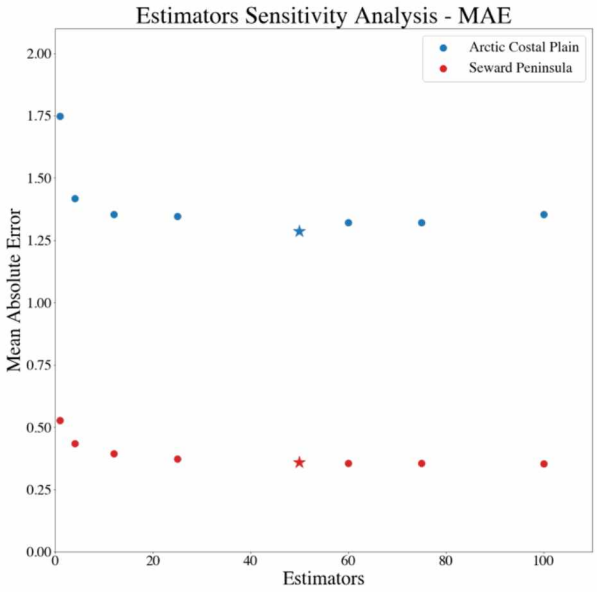

Figure 22: MAE for estimator sensitivity analysis

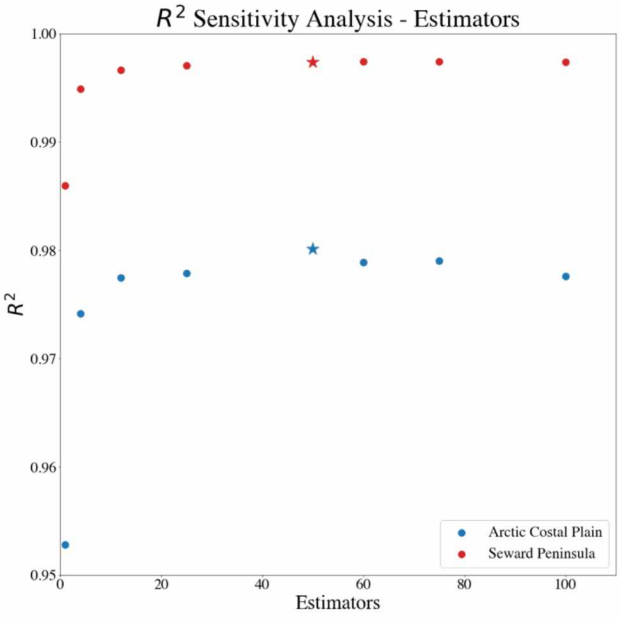

Figure 23: R^2 for estimator sensitivity analysis

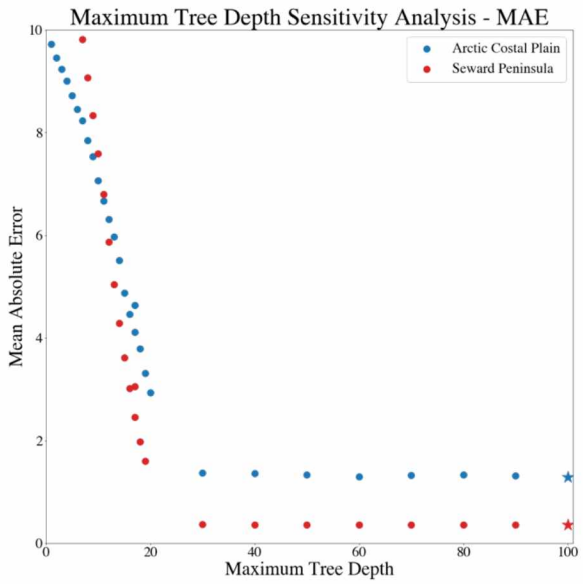

Figure 24: MAE for maximum tree depth sensitivity analysis

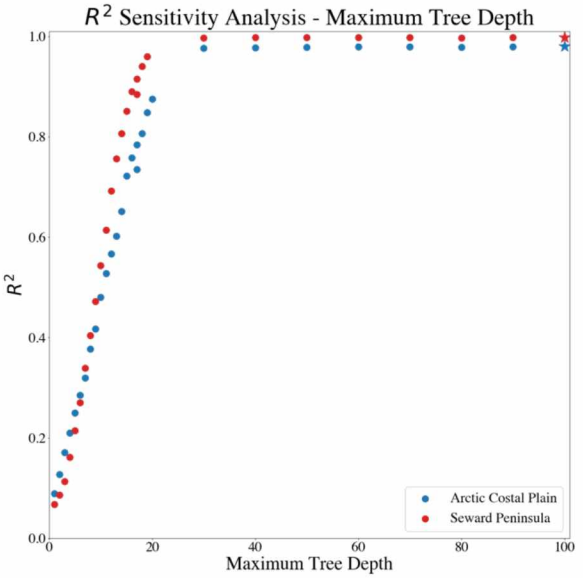

Figure 25: R^2 for maximum tree depth sensitivity analysis

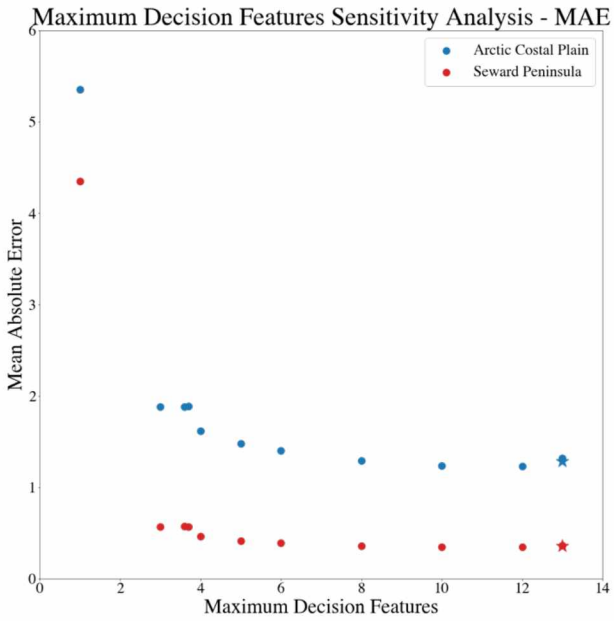

Figure 26: MAE for maximum decision feature sensitivity analysis

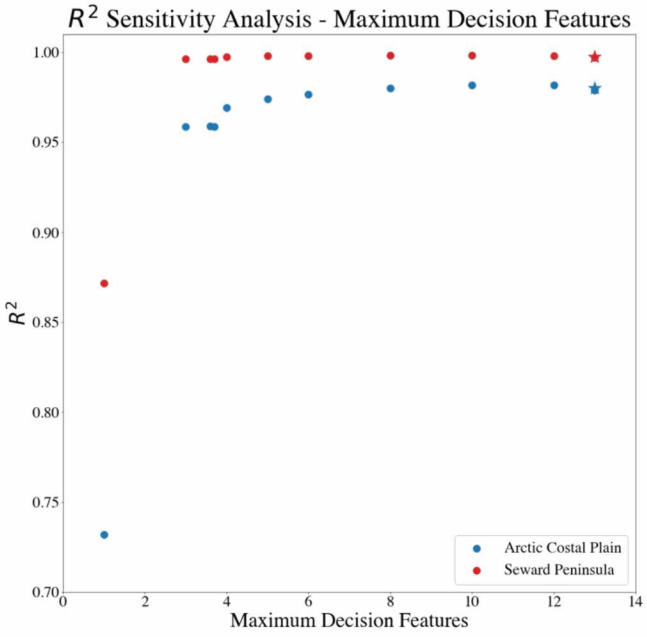

Figure 27: R^2 for maximum decision feature sensitivity analysis

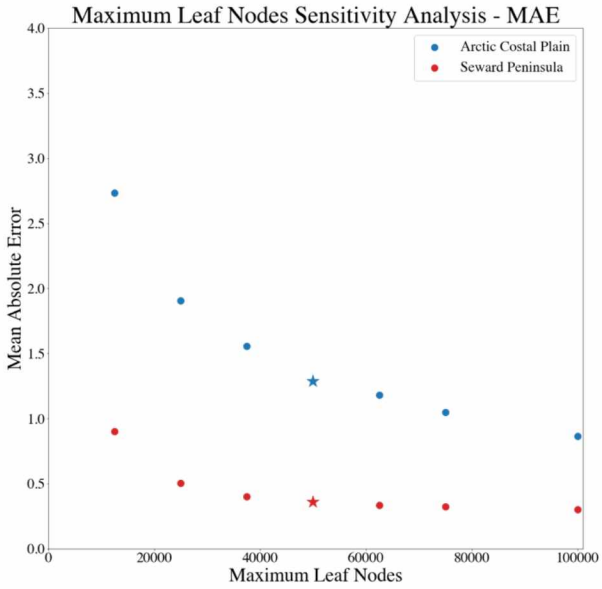

Figure 28: MAE for maximum leaf nodes sensitivity analysis

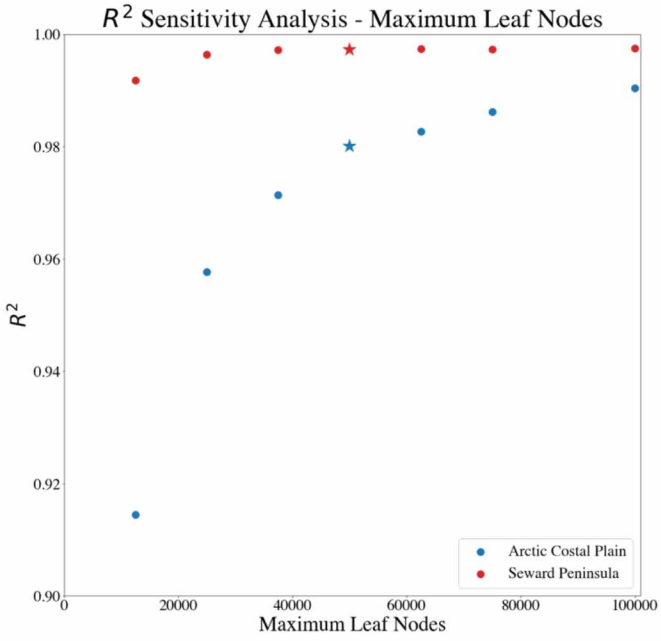

Figure 29: R^2 for maximum leaf nodes sensitivity analysis

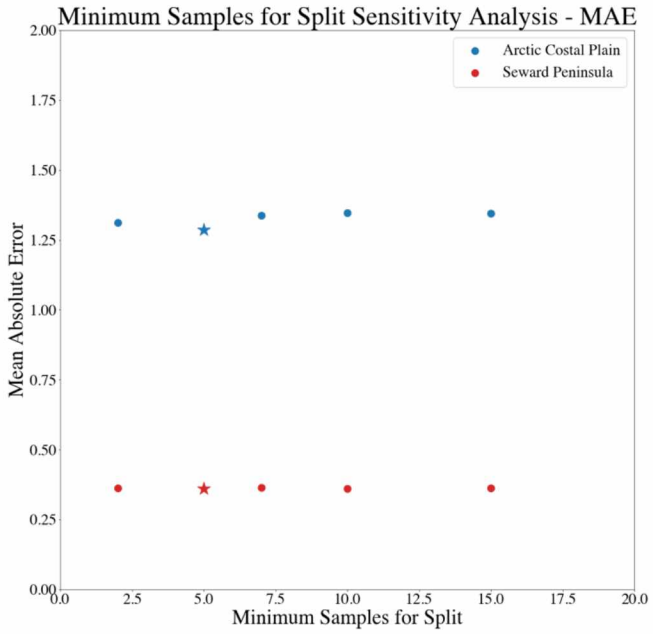

Figure 30: MAE for minimum samples for node split sensitivity analysis

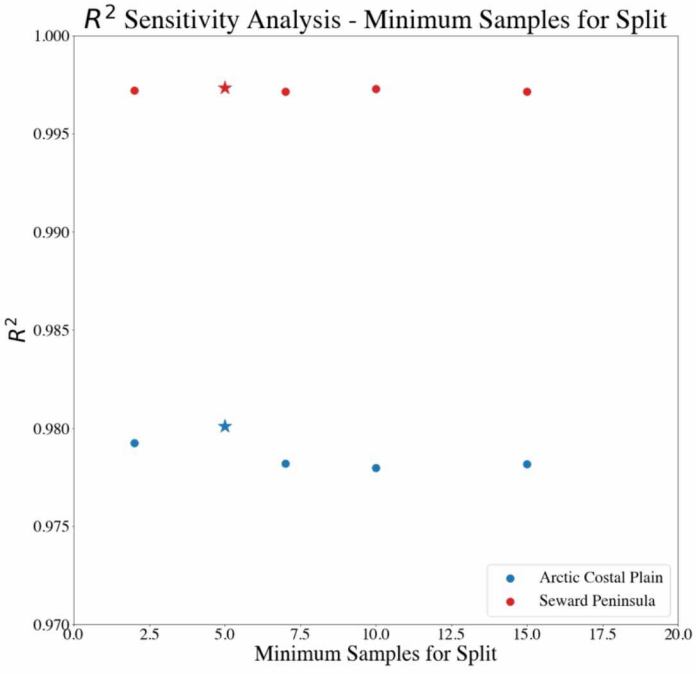

Figure 31: R^2 for minimum samples for node split sensitivity analysis

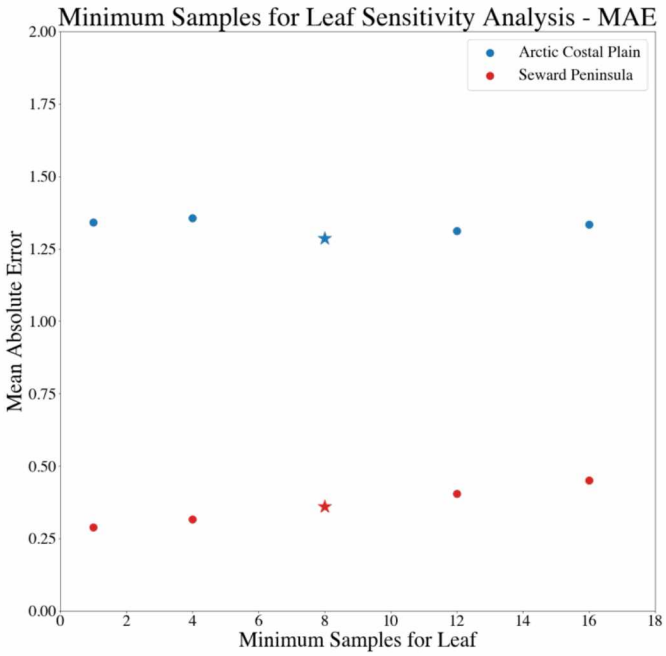

Figure 32: MAE for minimum samples for leaf sensitivity analysis

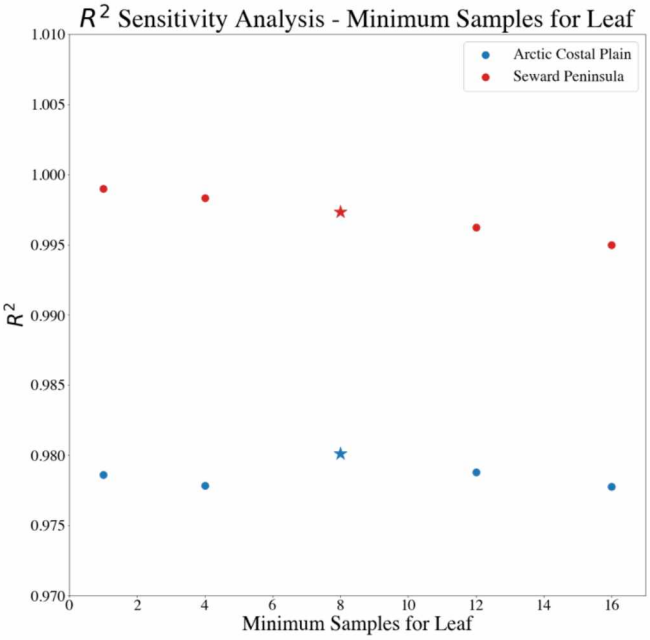

Figure 33: R^2 for minimum samples for leaf sensitivity analysis

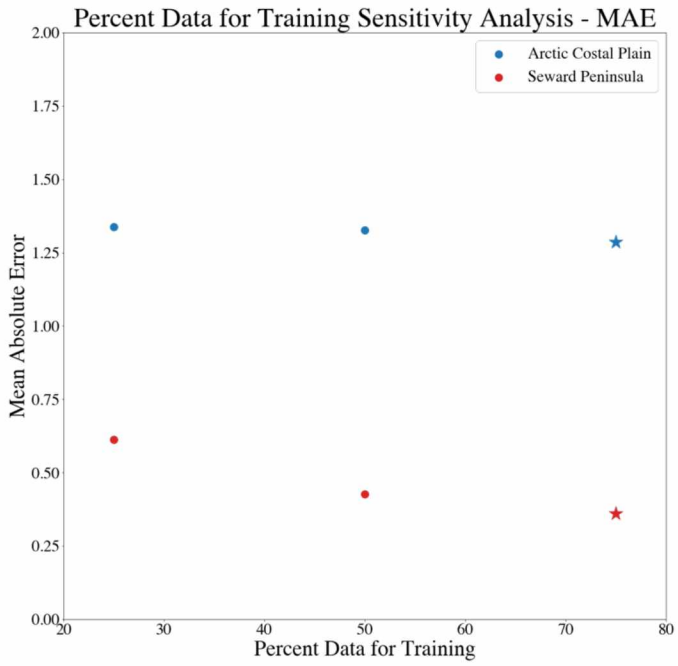

Figure 34: MAE for training data percent split sensitivity analysis

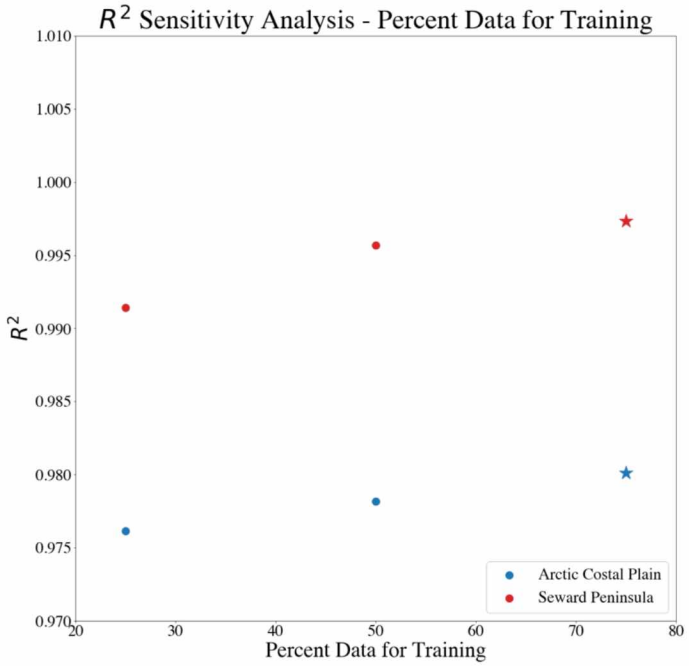

Figure 35: R^2 for training data percent split sensitivity analysis